Null Cones in Lorentz-Covariant General Relativity*

J. Brian Pitts†‡and W. C. Schieve§

November 1, 2001

Abstract

The oft-neglected issue of the causal structure in the flat spacetime approach to Einstein's theory of gravity is considered. Consistency requires that the flat metric's null cone be respected, but this does not automatically happen. After reviewing the history of this problem, we introduce a generalized eigenvector formalism to give a kine-

---

*This work is based on a portion of the first author's dissertation, supervised by the second author.

†The Ilya Prigogine Center for Studies in Statistical Mechanics and Complex Systems, The University of Texas at Austin, Austin, TX 78712 USA; present address: Mathematics Department, St. Edward's University, Austin, TX 78704 USA; email jpitts@physics.utexas.edu

‡J.B.P. thanks R. Hayes for making the null cone figure, I. Goldman for providing results from his unpublished dissertation in Hebrew, and M. Choptuik, L. Shepley, M. Visser, S. Deser, and A. A. Logunov for discussions. Support from the Ilya Prigogine Center for Studies in Statistical Mechanics and Complex Systems and the Welch Foundation is gratefully acknowledged.

§The Ilya Prigogine Center for Studies in Statistical Mechanics and Complex Systems, The University of Texas at Austin Austin, TX 78712 USA

matic description of the relation between the two null cones, based on the Segré classification of symmetric rank 2 tensors with respect to a Lorentzian metric. Then we propose a method to enforce special relativistic causality by using the gauge freedom to restrict the configuration space suitably. A set of new variables just covers this smaller configuration space and respects the flat metric's null cone automatically. Respecting the flat metric's null cone ensures that the spacetime is globally hyperbolic, indicating that the Hawking black hole information loss paradox does not arise in the special relativistic approach to Einstein's theory.

keywords: bimetric, causality principle, null cones, black hole information

suggested running head: Null Cones in Lorentz-Covariant General Relativity

1 Introduction

A number of authors have discussed the utility of a flat background metric $\eta_{\mu\nu}$ in general relativity or the possibility of deriving that theory, approximately or exactly, from a flat spacetime theory [1–115, 118, 120]. Some have permitted the background metric to be curved [70, 83, 95, 96, 121–125], but our interest is in flat backgrounds only, because they are uniquely plausible as nondynamical entities. The use of a flat background metric enables one to formulate a gravitational stress-energy tensor [94], not merely a pseudotensor, so gravitational energy and momentum are localized in a coordinate-independent (but gauge-variant) way. It also enables one to derive general relativity and other generally covariant theories from plausible special-relativistic postulates, rather than postulating them [13–18, 25–28, 30–33, 36, 37, 39–41, 43, 45–59, 66–70, 78–80, 82–86, 89–92, 107, 108, 112–115]. It is worth recalling a conclusion of E. R. Huggins [39], who was a student of Feynman. Huggins found that the requirement that energy be a spin-two field coupled to the stress-energy tensor does not lead to a unique theory, because of superpotential-type terms. Rather, "an additional restriction is necessary. For Feynman this restriction was that the equations of motion be obtained from an action principle; Einstein required that the gravitational field have a geometrical interpretation. Feynman showed these two restrictions to be equivalent." [39] (p. 3) Because other derivations have built in the requirement of an action principle already, it is no surprise that Riemannian geometrical theories are the unique result. (As W. Thirring observed, it is not clear *a priori* why Riemannian geometry is to be preferred over all the other sorts of geometry that exist, so a derivation of effective Riemannian geometry is attractive [26]. Casting gravitation in the same form as the other forces provides another reason to consider a field approach [26]. We will also find that the flat metric's null cone this approach avoids the difficulty of the lack of an *a priori* causal structure for defining dynamics in quantum gravity. Concerning general relativity, the lack of an *a priori* fixed causal structure is merely technically demanding at the classical level, but it constitutes a real puzzle at the quantum level, for one no longer knows how to write equal-time commutation relations, for example, because one needs to know the metric in order to determine equal times, but the metric is itself quantized: a chicken-and-the-egg problem.

Such derivations of general relativity and related theories, however, are only *formally* special relativistic, because the curved null cone might not respect the flat one. This difficulty afflicts not merely our derivation, but in fact all derivations in this tradition, and implies that the alleged resemblance of Einstein's theory to other field theories in this approach is merely formal, for all that has been shown to date. We survey in some detail the treatment of this fundamental question over the last six decades. As it happens, this issue has in general been ignored, explained away, postponed with the hope that it would go away, or mishandled in one of several ways, although there have been positive signs in recent years. We critique claims that the problem is insoluble and claims that it has already been solved, and conclude that the issue remains quite open.

Next we undertake to solve the problem. The kinematic issue of the relationship between the two null cones is handled using the work of G. S. Hall and collaborators on the Segré classification of symmetric rank 2 tensors with respect to a Lorentzian metric. For our purposes, we classify the curved metric with respect to the flat one, and find necessary and sufficient conditions for a suitable relationship.[1] Requiring that flat spacetime causality

[1] A portion of the mathematical work will appear in the *Proceedings of the 20th Texas Symposium on Relativistic Astrophysics*[119], and is used by permission of the American Institute of Physics.

not be violated, and not be arbitrarily close to being violated, a condition that we call "stable η-causality", implies that all suitable curved metrics have a complete set of generalized eigenvectors with respect to the flat metric, and that the causality conditions take the form of *strict* rather than loose inequalities. Given strict inequalities, one is in a position to solve such conditions, which are somewhat analogous to the "positivity conditions" of canonical gravity, which have been discussed by J. Klauder, F. Klotz, and J. Goldberg. In these new variables, stable η-causality holds *identically*, because the configuration space has been reduced, essentially by reducing the lapse until the proper null cone relation holds.[2] (The smaller space has the same dimension as the larger one, however.) This reduction implies the need for reconsidering the gauge freedom of the theory. It turns out that gauge transformations no longer form a group, because multiplication is not defined between some elements.

Making the curved metric respect the flat null cone ensures that the resulting spacetime is *globally hyperbolic*. This fact is quite consistent with the existence of a region of no escape, which can arise due to the inward tilting of the curved null cones [98]. Given that global hyperbolicity apparently pulls the fangs from the Hawking black hole information loss paradox, it appears that this paradox does not afflict the special relativistic approach to Einstein's theory of gravitation. This lack of paradox indicates that the

[2]This approach might have the consequence that the lapse is forced toward 0 in some contexts. In such cases, it might turn out that objectionable features are simply removed from Minkowski spacetime by infinite postponement.

flat metric can be employed such that it is not merely a formal mathematical trick, but rather has beneficial physical consequences.

2 Bimetric General Relativity and Null Cone Consistency: A History Since 1939

As we have seen, the use of a flat metric tensor $\eta_{\mu\nu}$ in gravitation has received a fair amount of attention over the last six decades or so. However, the interpretation of the resulting bimetric or field formulation of general relativity has not been adequately clarified, due to an ambiguous notion of causality: the effective curved metric which determines matter propagation is not obviously consistent with the flat background causal structure. Having a consistent relationship is clearly a *necessary* condition for a true special-relativistic theory.[3] Examples of consistent and inconsistent relationships are given in figure 1.

Whether it is *sufficient* is unclear from anything said so far, because the propagation of gravity itself or of nonminimally coupled matter fields can yield more subtle behavior [151, 154]. However, we will see below that this

[3]This question arises for slightly bimetric theories [112] and for massive versions of general relativity as well, but our solution of reducing the lapse to solve the problem might not be available for slightly bimetric theories, and definitely is unavailable for massive theories, at least in their usual form. If a procedure such as BFT could be used to turn the second-class constraints of a massive theory into first-class constraints *without* introducing nonlocality due to the infinite order derivatives present in finite gauge transformations, then massive theories would also admit our procedure for ensuring null cone consistency.

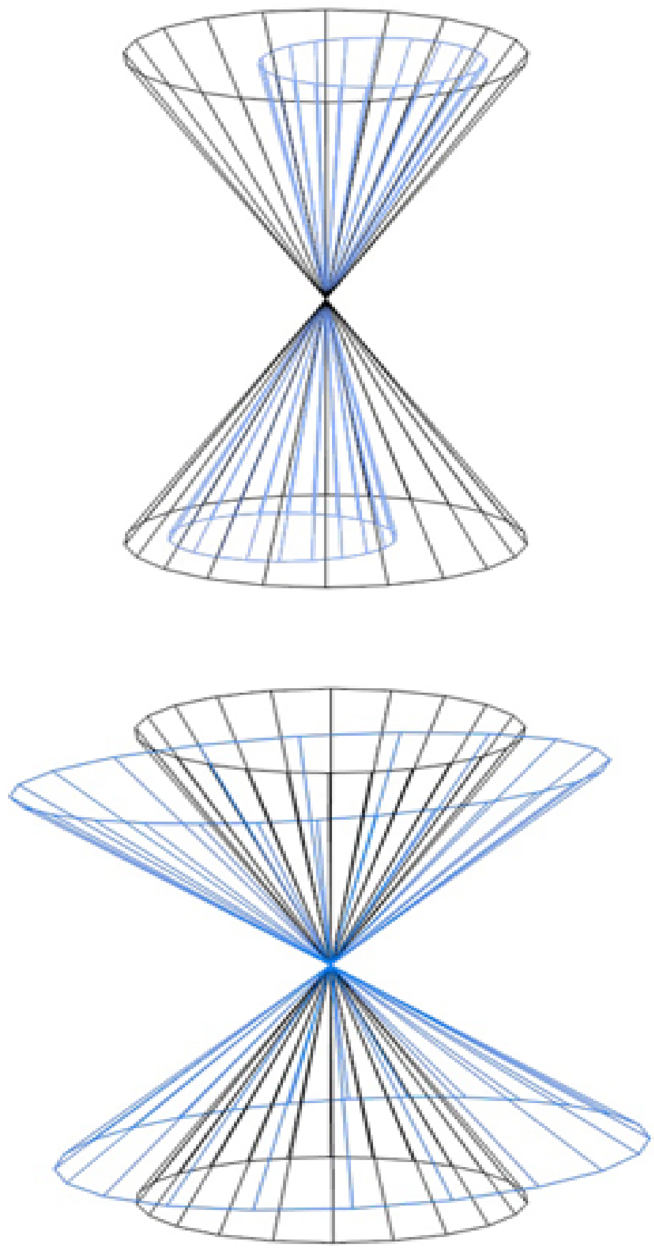

Figure 1: Examples of Consistent (above) and Inconsistent (below) Null Cone Relationships in 2 Dimensions. By Ryan Hayes using *Blender*

null cone condition indeed is sufficient due to the presence of a well posed initial value formulation of the special relativistic approach to Einstein's equations.

We now sketch the history of the flat null cone issue from roughly the late 1930s till the present. We do not consider the period between 1905 and the late 1930s, though doubtless that would be an interesting project, which would consider the time after special relativity had solidified and include the invention of Einstein's theory of gravitation. (L. P. Grishchuk mentions a bit of the history in works of Poincaré and Einstein [86]. Fang and Fronsdal sketch the history of the flat spacetime approach up to 1979 [79], but neglect to consider the null cone issue.) Rather, we start with the *rebirth* of the flat spacetime approach to gravity, with works by Fierz, Rosen *et al.* While the importance of the problem perhaps seems evident in retrospect, the neglect of it in the literature suggests that it in practice was not so obvious, or that influential radical empiricist philosophies obscured its importance. One can crudely divide the issue's history into three periods, though at times we will disregard the historical boundaries to be able to discuss an author's whole work in a unified way. For the first 20 years (1939-1959), the problem seems not to have been recognized or mentioned, at least not in print (to our knowledge). For the next two decades (1959-1979), it was sometimes mentioned, but either resolved incorrectly, dismissed as unimportant, or postponed with the hope that it would disappear. More recently (1979-2001), it has been recognized more often, and occasionally regarded as worthy of sustained attention. A few authors have attempted either to solve it or to prove it insoluble. However, we disagree that either of these goals has been achieved, and will undertake to show why.

One should perhaps distinguish between two null cone problems. The first is: given that one regards Einstein's equations as describing the evolution of an effective curved metric in Minkowski spacetime, what does one make of the potential violation of Minkowski causality by matter responding to the curved metric? The second is: given that one chooses to quantize the geometrical (single-metric) theory, what does one make of causality without a metric to define equal-time commutation relations? However, these problems are related, and we believe that the SRA as presented here solves them both, so we will treat them together.

2.1 The Years 1939 to 1959: the Null Cone Consistency Problem Ignored

Around 1940, in his seminal papers on the bimetric description of general relativity, N. Rosen suggested that there ought to be some (gauge-fixing) relation between the flat and curved metrics, because one expects that the two coincide if the gravitational field vanishes [3] (p. 149). While this paper did not consider the meaning of the bimetric formalism in detail, its companion paper (p. 150) considered interpretive issues. Rosen wrote (apart from a change in notation to match ours),

> [f]rom the standpoint of the general theory of relativity, one must

> look upon $\eta_{\mu\nu}$ as a fiction introduced for mathematical convenience. However, the question arises whether it may not be possible to adopt a different point of view, one in which the metric tensor $\eta_{\mu\nu}$ is given a real physical significance as describing the geometrical properties of space, which is therefore taken to be flat, whereas the tensor $g_{\mu\nu}$ is to be regarded as describing the gravitational field. [3] (p. 150).

Rosen recognized that the flat spacetime view implies that the speed of light (measured with ideal rods and clocks, which are not distorted by gravity) will tend to differ from unity [3] (p. 153), but he seems not to have addressed the possibility that it might *exceed* 1. While his approach merely postulated bimetric general relativity, he did suggest that it would be desirable to derive it independently [3] (p. 153). His intention to carry out this procedure himself [3] (p. 153) seems not to have been realized, but many others have done it since that time, as did we in an earlier chapter.

During the 1940s, with some war-time inconvenience in Greece, A. Papapetrou was able to express general relativity in an attractive form resembling electromagnetism, with the theories being expressed in the tensorial DeDonder and Lorentz gauges, respectively [8]. He emphasized the improved nature of the conservation laws, especially for angular momentum, and found that certain attractive relations that have no invariant meaning in the geometrical view become perspicuous given the flat spacetime interpretation. Papapetrou held that for the flat spacetime approach, gauge-fixing to tie together the two metrics was "indispensable" (p. 20), because the energy-momentum and angular momentum localization would suggest physically distinct systems given different relations between the two metrics. He was aware of Rosen's result that the flat spacetime interpretation implies a varying speed of light (using unrenormalized instruments), but seems also to have failed to entertain the possibility that the gravitational field might make light travel *faster* than in special relativity.

The neglect of the null cone issue continued well into the 1950s in the important works of S. N. Gupta [13–16] and R. H. Kraichnan [17, 18]. At this stage the derivation of the exact nonlinearities of general relativity, which Rosen had desired, was achieved. Concerning the special-relativistic nature of the theory, both authors seem to have regarded the Lorentz covariance of the theory as sufficient for special relativity. If the theory's gauge invariance and the unobservability of the flat metric are mentioned, the idea that the observable effective curved metric might well *conflict* with the flat metric is not. This is an important distinction that will also be overlooked repeatedly by later authors. One could imagine that the flat metric might fail to appear in the equations of motion, but still have its null cone serve as a bound on the curved metric's null cone, so this distinction is not trivial. The flat metric might have important qualitative consequences, even without having any quantitative role in the field equations.

F. J. Belinfante, interested in the work of Papapetrou and Gupta and in particular in the solidifying covariant perturbation approach to quantizing gravity, contemplated the use of a flat metric in "Einstein's curved universe", which evidently meant the geometrical theory of gravity [19]. Working in the context of the static Schwarzschild solution (in which it is difficult to get the relation between the two null cones wrong, at least outside the Schwarzschild radius, unless one tries to do so), Belinfante only had occasion to consider the null cone relationship incidentally. But the fact that r becomes g-temporal and t becomes g-spatial for very small radii, juxtaposed with the *a priori* fixed character of these quantities with respect to the flat metric, does give him occasion for thought. Belinfante gives indications (including in the paper's title) that he does not believe deeply in the flat spacetime approach, so perhaps the null cone issue would not have interested him. While he is prepared to suggest that the "Swiss-cheese"-like behavior of the Schwarzschild solution in the bimetric context might help eliminate field theory's divergences, it is clear from review papers on quantum gravity [22, 23] that the flat metric is just a tool–perhaps a useful one, but more likely not–for Belinfante. It is thus not too surprising that the null cone issue is ignored. The "[r]eal problem" is not to be found in "[t]heories, usually in flat space, which seek to be approximations to Einstein's theory, or a perturbation-theoretical treatment of Einstein's theory", but in "[q]uantization of Einstein's theory itself." (pp. 198, 192) [22]. For Belinfante, spacetime might have a Swiss cheese structure, contain worm holes, or have a closed spatial topology [23]. Some of Belinfante's work with Swihart on linear gravity also neglects to discuss the null cone issue [20, 21].

2.2 The Years 1959-1979: the Problem Dismissed or Postponed

The null cone consistency issue is perhaps first discussed in print by W. Thirring in 1959 [25–27], but then dismissed with a resolution that does not permit a true special relativistic interpretation. Thirring clearly recognizes the apparent conflict between the two null cones [26], writing, "Another feature of the equations of motion ... we want to point out is that the velocity $|d\mathbf{x}/dt|$ is not required to be < 1 ... Thus [assuming the curved metric to be diagonal] there is a limiting velocity c but it is space dependent and may exceed unity." (pp. 100, 101) A bit later, he writes "Since c is also gauge dependent and will exceed unity in some gauge systems [the matter equation of motion] even admits an apparently acausal behavior." (p. 101) However, Thirring thinks that this acausal behavior is *only* apparent, for he is satisfied with the fact that the "renormalized" velocity (measured physically using real clocks, which are distorted by the gravitational field) is not greater than unity: "However, we shall see shortly that c also corresponds to the velocity of light and that it becomes unity when measured with real measuring rods and clocks since they all are affected by the [gravitational] field." (p. 101) Evidently the unobservable nature of the intervals governed by $\eta_{\mu\nu}$ satisfies Thirring that the apparently acausal behavior is not a problem: "The real metric [interval corresponding to $g_{\mu\nu}$] is gauge invariant whereas [the interval corresponding to $\eta_{\mu\nu}$] is not and therefore has no physical significance. Space-time measured with real objects will show a Riemannian structure whereas there are no measuring rods which could measure the original pseudoeuclidean space." (p. 103) Thirring's argument is doubtful because the same distinction that was neglected by Gupta and Kraichnan is also neglected here: the non-measurability of the flat metric does not entail that it lacks physical significance. Generally one considers causality to be an important physical concept. At the risk of stating the obvious, we recall that in special relativity, the relevant speed for causality is not the speed at which electromagnetic radiation actually propagates, but the value of the universal velocity constant (ordinary called "the speed of light" and written as c, but to do so here would invite confusion) which appears in Lorentz transformations, that is essential. As is well-known, to permit propagation faster than that speed in one frame is to admit backward causation–which is usually rejected–in another frame. Given the violation of the flat spacetime null cone, it is not clear what Thirring's field theoretic approach means. Yet, according to Thirring, the field theoretic approach gives "a theory following the pattern of well understood field theories, in particular electrodynamics." (p. 116) Thus, Thirring's list of advantages and disadvantages of the field and geometric approaches to gravity (pp. 116, 117) is notably incomplete, because the obvious notion of causality for the field approach has been discarded. Thirring comes very close to noticing the problem of null cone inconsistency, but then stops short, apparently due to a prejudice against unobservable entities.

One might hope that Thirring's almost-recognition of the problem would have inspired his successors to recognize and perhaps try to solve it. That, however, did not occur. In particular, although L. Halpern made a rather minute study of Thirring's paper [37], the light cone issue receives only a single sentence (p. 388), one sufficiently noncommittal that no discomfort with Thirring's purported resolution of the causality issue is obvious. Halpern was not an advocate of the flat spacetime approach to gravitation [36], so it is the more remarkable that he overlooked a potentially serious difficulty. R. Sexl also was aware of the Thirring's work and even presented it at a conference [30], yet he also accepted Thirring's ostensible resolution of the null cone conflict [30, 31].

The covariant perturbation program for quantizing general relativity yielded a large number of works based on expanding the curved metric into a background part and a dynamical part. Thus, one might expect the question of the relation of the two metrics to be considered in some way. Commonly the background metric was flat, leading to equations at least formally special relativistic.

A notable exception is the work of B. S. DeWitt, who made great use of non-flat background metrics and found various benefits in doing so [122]. While DeWitt could make use of a background metric, to him it was always at most a tool, not a deep part of nature. In an article entitled "The Quantization of Geometry," he wrote:

> The problem of [quantizing the gravitational field] may be approached from either of two viewpoints, loosely described as the

> "flat space-time approach" and "the geometrical approach." In the flat spacetime approach, which has been investigated by several authors ... the gravitational field is regarded as just one of several known physical fields, describable within the Lorentz-invariant framework of a flat space-time. Its couplings with other fields ... lead to a contraction or elongation of "rigid" rods and a retardation or advancement of "standard" clocks .... Both the geometrical and flat space-time points of view have the same *real* physical content. However, it has been argued that the flat space-time approach provides more immediate access to the concepts of conventional quantum field theory and allows the techniques of the latter to be directly applied to gravitation. While there is merit in this argument, too strong an insistence upon it would constitute a failure to have learned the lessons which special relativity itself has already taught. Just as it is now universally recognized as inconvenient (although *possible*) to regard the Lorentz-Fitzgerald contraction from relativistic modifications in the force law between atoms, so it will almost certainly prove inconvenient at some stage to approach space-time geometry, even in the quantum domain, in terms of fluctuations of standard intervals which are the same for all physical devices and hence unobservable. [121] (pp. 267,268).

Concerning the question of a well-defined causal structure, which his approach appeared to lack, he later suggested, "Critics of the program to quantize gravity frequency [*sic*] ask, 'What can this mean?' A good answer to this question does not yet exist. However, there are some indications where the answer may lie." [122]. DeWitt's vision for the program, which he was prepared to call "covariant quantum geometrodynamics" in a volume honoring J. A. Wheeler (the title itself suggesting sympathy for a geometrical view of gravitation, much as "The Quantization of Geometry" did), included that it "should be able to handle any topology which may be imposed on 3-space"[123] (p. 437). Of the covariant perturbation formalism, he wrote that the "most serious present defect of the covariant formalism is its foundation in scattering theory, with spacetime being assumed asymptotically flat. The method of the background field, which we have introduced, indicates a way in which this defect may be removed" [123] (p. 437). It seems very likely that the null cone issue, as we have formulated it, would not be important to DeWitt, given that the background metric was merely a tool for investigating a truly geometrical theory.

Other authors, especially in the particle physics tradition, seem at least somewhat more content with a flat background metric. In his lectures on gravitation, R. Feynman shows himself ambivalent about the interpretation of gravitation. After deriving Einstein's equations from a flat spacetime field theory, he concluded from the unobservability of the flat metric that the latter was not essential. Using an analogy with curiously intelligent insects walking on a tiled floor, he says that "[t]here is no need to think of processes as occurring in a space which is truly Euclidean, since there is nothing physical which can ever be measured in this fictional space. The tiles represent simply a labelling of coordinates, and any other labelling would have done just as well" [40] (p. 101). Concerning the "assumption that space is truly flat," he concludes that "[i]t may be convenient in order to write a theory in the beginning to assume that measurements are made in a space that is in principle Galilean, but after we get through predicting real effects, we see that the Galilean space has no significance" (p. 112), but serves only as a "bookkeeping device" (p. 113). Concerning the "relations between different approaches to gravity theory,"

> [i]t is one of the peculiar aspects of the theory of gravitation, that it has both a field interpretation and a geometrical interpretation . . . these are truly two aspects of the same theory . . . the fact is that a spin-two field has this geometrical interpretation; this is not something readily explainable–it is just marvelous. The geometric interpretation is not really necessary or essential to physics. It might be that the whole coincidence might be understood as representing some kind of gauge invariance. It might be that the relationship between these two points of view about gravity might be transparent after we discuss a third point of view . . . . (p. 113)

Feynman seems to feel free to switch between the two views as he sees fit. Questions about nontrivial topologies or the desire to have a transparent notion of causality seem not to have occupied him. Had they, he might have hesitated in proclaiming them to be "the same theory", given the competition between causality and gauge invariance. Later, in developing the covariant perturbation theory, Feynman did not address these issues, but wrote as if no conceptual difficulties existed. He wrote: "The questions about making a 'quantum theory of geometry' or other conceptual questions are all evaded by considering the gravitational field as just a spin-2 field nonlinearly coupled to matter and itself (one way, for example, is expanding $g_{\mu\nu} = \delta_{\mu\nu} + h_{\mu\nu}$ and considering $h_{\mu\nu}$ as the field variable) and attempting to quantize this by following the prescription of quantum field theory, as one expects to do with any other field. The central difficulty springs from the fact that the Lagrangian is invariant under a gauge group," but this issue, he finds, can be resolved by adding a gauge-fixing term, the result being "completely satisfactory" at the level of tree diagrams (which correspond to the classical theory) [42]; see also ([41]). If the main difficulty is gauge invariance–which in fact competes with special relativistic causality, as Thirring nearly realized–and if the gauge-fixing terms lead to a "completely satisfactory" result without regard to the light cone relationship, then, unless we are to charge Feynman with oversight, clearly the flat metric is merely a useful tool for him. However, the claim to have avoided all conceptual questions cannot be sustained, because the light cone issue is just such a question, and the meaning of parts of Lorentz-covariant field theory remains obscure if the problem is ignored. Huggins, a student of

Feynman, also neglects to consider the null cones issue [39].

S. Mandelstam presented a critique of the flat-space covariant perturbation program as Gupta had developed it [34]. Gupta had imposed the DeDonder coordinate (gauge) condition. Let us see how close Mandelstam comes to identifying the null cones issue. He writes: “Quantization in flat space can only be regarded as a provisional solution of the problem for several reasons,” such as its approximate (at least at that stage of development) character, the use of an indefinite metric, and the presence of unphysical states. “But the main objection to this method of quantization lies surely in the physical sacrifices it makes by going to flat space. The variable specifying the coordinates are numbers without physical significance which can be chosen in an infinite variety of ways. On the other hand, distances in space-time, which are physically significant entities, are related to the coordinates in a manner which has not been elucidated when the metric is quantized.” However, perhaps these objections can be met: “It may be possible to add to the theory a prescription for interpreting its results physically. If it could then be shown that the predictions of the theory were independent of the coordinate conditions used, and that they tended to the predictions of the unquantized theory in the classical limit, we would have a satisfactory theory. Some progress has actually been made in this direction by Thirring”, which “indicates the connection of the Gupta variables to the metric,” though “the basic difficulties of the ‘flat space’ approach remain.” Clearly one of Mandelstam’s worries is the question of gauge invariance in a procedure that makes use of coordinate conditions. It is difficult to tease out a clear statement of worry about rival null cones from these remarks, though the issue might have been intended among the "the basic difficulties of the 'flat space' approach" that remain.

A moment of considerable clarity occurred in 1962 with the appearance of a paper by J. R. Klauder [38], whose abstract opens with the statement, "[i]n any quantum theory, in which the metric tensor of Einstein's gravitational theory is also quantized, it becomes meaningless to ask for an initial space-like surface on which to specify the conventional field commutators." Klauder elaborates:

> In so far as [certain] formalisms [for quantizing gravity] are transcriptions of techniques successful in a flat Lorentz space-time, they ignore a unique problem peculiar to general relativity. Conventional field theories deal, in particular, with commutation rules, which, when employed for the fields separated by a space-like interval, have an especially simple form. Whether two nearby points are or are not space-like is a *metric* question that can be asked (and in principle answered) not only in flat space but also in any space with a preassigned curved metric as well. However as soon as the space-time metric $g_{\mu\nu}(x)$ becomes a dynamical variable–as in Einstein's theory–then an initial space-like surface on which to specify commutators of any two fields becomes a meaningless concept.

Klauder's approach to handling this problem was to propose an alternative formalism in which fields can fail to commute at most only at the same *event.* Unfortunately, Klauder's acute awareness of the null cone issue did not spread widely.

S. Weinberg did considerable work on gravitation considered as a Lorentz-invariant theory [45–53]. Concerning the geometric interpretation of general relativity, Weinberg could write that "the geometric interpretation of the theory of gravitation has dwindled to a mere analogy, which lingers in our language ... but is not otherwise very useful. The important thing is to be able to make predictions about images on photographic plates, frequencies of spectral lines, and so on, and it simply doesn't matter whether we ascribe these predictions to the physical effect of gravitational fields or to a curvature of space and time." [52] (p. 147) This ambivalence about the meaning of the theory perhaps helps to explain why the null cone consistency issue appears to be ignored in Weinberg's writings. However, the meaning of concepts used in Lorentz-invariant field theory in which Weinberg's work is rooted, or at least its relation to an underlying classical theory, does seem somewhat obscure if this issue is neglected. Somewhat more recently, R. Penrose reported that Weinberg was "no longer convinced that the anti-geometrical viewpoint is necessarily the most fruitful" [54], on account of some impossibility theorems [53]. In recent personal communication with one of us (J. B. P.), Weinberg stated that he is no longer a strong advocate of any view on the subject, though it is quite interesting that the flat spacetime approach reproduces general relativity.

Based on the "spin limitation principle," which requires that only definite angular momenta be exchanged, V. I. Ogievetsky and I. V. Polubarinov have derived Einstein's equations and a family of massive relatives thereof in flat spacetime [57–59]. While this principle is quite attractive, it fails to pay any heed to whether the resulting theories yield propagation consistent with the causal structure of the flat metric. Given that some of their theories are massive and thus make the $\eta_{\mu\nu}$ *observable*, this shortcoming seems fairly serious. While we can find no mention of the null cone issue in the work of Ogievetksy and Polubarinov, it would be interesting to see if the spin limitation principle could be generalized in such a way as to yield consistency of the null cones.

A large amount of work related to the field approach has been done by S. Deser, sometimes with collaborators such as D. G. Boulware, R. Nepomechie, A. Waldron or others. In the course of papers which derived general relativity via self-interaction in flat spacetime [66] or curved [70], or general relativity from quantum gravity [67, 68], or supergravity from self-interaction [69], or which study bimetric theories for a festschrift for N. Rosen [73], we can find no mention of the issue of the null cone consistency issue. In particular, Deser finds the main issues for bimetric theories to be essentially the same problems that he and Boulware found in massive variants of general relativity [153], *viz.*, empirically falsified light bending properties, negative energy disasters, or both [73]. Unlike some authors who have a strong preference, Deser (after a rather pro-geometrical paper early on [155]) seems to admire both the geometric and field formulations: "The beautiful geometrical significance of general relativity is complemented by its alternate formulation as the unique consistent self-coupled theory arising from flat-space free gravitons, without appeal to general covariance." [70] However, depending on how one reads the flat spacetime approach, one might obtain some different features, as we will observe below, so one might prefer to see the meaning of the field formalism addressed. Deser and R. Nepomechie have studied a somewhat related issue related to the anomalous propagation of gauge fields in some conformally flat spacetimes, compared to a flat background [71, 72] with the same null cone structure. In particular, backscattering off the geometry causes the propagation to lie not merely on the null cone, but inside it. However, "while our results are surprising, they do not imply any consistency problems" [71], as they would if the propagation were *outside* the null cone.[4]

Some time ago it was recognized by G. Velo and D. Zwanziger that spin $\frac{3}{2}$ field propagation has causality worries. They found that the "main lesson to be drawn from our analysis is that special relativity is not automatically satisfied by writing equations that transform covariantly. In addition, the solutions must not propagate faster than light" [152]. A rather similar lesson needs to be learned regarding the gravitational field (spin 2) as well, for this is just the point that Gupta and Kraichnan missed.

During the 1970s, the relation between the two null cones continued to be

[4]We thank Prof. Deser for calling our attention to this issue and the spin $\frac{3}{2}$ field.

neglected in the context of the covariant perturbation approach to quantizing general relativity, at least in practice. However, quantum gravity review talks drew attention to this problem from time to time. This service was performed with special clarity in a 1973 review by A. Ashtekar and R. Geroch [127]. They find that much of the difficulty in quantizing the theory arises from the fact that "the distinction between the arena and the phenomenon, characteristic of other physical theories, is simply not available in general relativity: the metric plays both roles." [127] (p. 1214) In discussing field theoretic approaches, they write that

> [i]t is normally the case in quantum field theory ... that two distinct fields come into play–a kinematical background field (the metric of Minkowski space) and a dynamical field .... One can certainly regard general relativity as a field theory, but in this case there is only a single field, the metric g_{ab} of spacetime, which must play both these roles. But the application of the techniques of quantum field theory apparently requires a non-dynamical background field. In quantum electrodynamics, for example, the causality of the Feynman propagators and the asymptotic states, in terms of which the S-matrix is defined, refer directly to the metric of Minkowski space. Thus, one does not expect to be able to carry over directly to general relativity, regarded as a classical field theory, the procedure which led for example from classical Maxwell theory to quantum electrodynamics. In order to apply

> the techniques of quantum field theory one must, apparently, either modify these techniques or reformulate the interpretation of general relativity as a field theory. (p. 1229)

This latter suggestion of reinterpreting general relativity does not strike us as being unimaginable or even excessively difficult, especially given that Kraichnan had presented a simple and clean derivation already in 1955 [17]. However, Ashtekar and Geroch do present some objections to this general line of attack. "It turns out, however, that this perturbation approach to obtaining a quantum theory of the gravitational field suffers from a number of difficulties. There exist, [128] for example, four-dimensional manifolds M on which there are metrics g_{ab} of Lorentz signature, but on which there are no flat metrics." (p. 1232) However, it is not clear why such examples must be regarded as physically admissible. If it could be shown that some exact solutions of obvious physical utility admit no flat metric, then the argument would be persuasive, but that argument was not made. Thus, it seems that this argument against the perturbation approach will be highly persuasive only if one is *already* committed to a geometrical view of Einstein's equations at the classical level. But one would exaggerate only slightly to say that that is the point at issue. The idea of *requiring* that the curved null cone be consistent with the flat one seems not to have been entertained, but Ashtekar and Geroch have shown powerfully, if reluctantly, why such an approach merits consideration.

The null cone consistency issue was emphasized by C. J. Isham at the first Oxford quantum gravity symposium [167]. We quote from pp. 20, 21:

> One natural approach perhaps is to separate out the Minkowski metric $\eta_{\mu\nu}$ and write $g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x)$ where $h_{\mu\nu}(x)$ describes the deviation of the geometry from flatness ... [which approach has some advantages.] However, there are a number of objections to this point of view. For example: (i) The actual background manifold may not be remotely Minkowskian in either its topological or metrical properties, in which case the separation [above] ... is completely inappropriate. (ii) Even if [the equation above] is justified (from the point of view of i)) [*sic*] the procedure is still dubious because the lightcone structure of the physical spacetime is different from that of Minkowski space. For example, if the field $\hat{\phi}$ has some sort of microcausality property with respect to the metric $g_{\mu\nu}$ then this is <u>not</u> equivalent to microcausality with respect to the fictitious Minkowski background.

Once again a prior commitment to a geometrical view of general relativity is manifest. Isham seems not to entertain the idea of *requiring* that the spacetime be compatible with the flat background, but at least the consistency issue is clearly stated. At the second Oxford symposium, Isham observed that "one of the ambitions of the Riemannian programme is to free quantum gravity from perturbation theory based on the expansion $g_{\mu\nu} = g^c_{\mu\nu} + \sqrt{G} h_{\mu\nu}$. Expansions of this type are known to be bad in classical general relativity and they clearly misrepresent the global topological and lightcone structures of the pair $(M, g_{\mu\nu})$" [168] (p. 14). Isham has continued to mention this issue in more recent talks in the context of the problem of causality and time [139, 179]. The problem of time shows up in the light cone issue for the covariant perturbation approach to quantum gravity, but related difficulties show up elsewhere [179]. Still more recently, Isham has expressed the issue as follows:

> *The problem of time* The background metric η provides a fixed causal structure with the associated family of Lorentzian inertial frames. Thus, at this level, there is no problem of time. The causal structure also allows a notion of microcausality, thereby permitting a conventional type of relativistic quantum field theory to be applied to the field $h_{\alpha\beta}$.
>
> However, many people object strongly to an expansion [of the curved metric into a flat one plus a dynamical part] since it is unclear how this background causal structure is to be related to the physical one; or, indeed, what the latter really means . . . it is not clear what happens to the microcausal commutativity conditions in such circumstances; or, indeed, what is meant in general by 'causality' and 'time' in a system whose light cones are themselves the subject of quantum fluctuations.[139] (p. 58)

Another moment of awareness of the null cone consistency issue in a quantum gravity review talk comes from P. van Nieuwenhuizen at the first

Marcel Grossmann meeting. After showing keen awareness of the problem, van Nieuwenhuizen shelves it. He writes:

> According to the particle physics approach, gravitons are treated on exactly the same basis as other particles such as photons and electrons. In particular, particles (including gravitons) are always in flat Minkowski space and move as if they followed their geodesics in curved spacetime because of the dynamics of multiple graviton exchange. This particle physics approach is entirely equivalent to the usual geometric approach. Pure relativists often become somewhat uneasy at this point because of the following two aspects entirely peculiar to gravitation:
>
> - In canonical quantization one must decide before quantization which points are spacelike separated and which are timelike separated, in order to define the basic commutation relations. However, it is only after quantization that the fully quantized metric field can tell us this spacetime structure. It follows that the concept of space-like or time-like separation has to be preserved under quantization, and it is not clear whether this is the case. (One might wonder whether the causal structure of spacetime need be the same in covariant quantization as in canonical quantization.)
> - Suppose one wanted to quantize the fluctuations (for example of a scalar field, or even of the gravitational field itself)

> about a given curved classical background instead of about flat Minkowski spacetime. In order to write the field operators corresponding to these fluctuations in second-quantized form, one needs positive and negative frequency (annihilation and creation) solutions. In non-stationary spacetimes it is not clear whether one can define such solutions. (It may help to think of non-stationary space-time as giving rise to a time-dependent Hamiltonian.)
>
> The strategy of particle physicists has been to ignore these two problems for the time being, in the hope that they will ultimately be resolved in the final theory. Consequently we will not discuss them any further.[76]

While quantization is not our immediate concern, a similar worry to the first of these two exists at the classical level if one wishes to take the flat metric seriously: there is no reason to expect that the dynamics will yield automatically a physical causal structure consistent with the *a priori* special-relativistic one, but inconsistency leads to grave interpretive difficulties.

2.3 The Years 1979-2001: the Problem Increasingly Attended and the Development of Three Views

More recently, the question of null cone consistency has come to be recognized as interesting somewhat more often. While a fair number continue to neglect the issue, those who have addressed it can be found to have one of three attitudes toward the flat metric: that it is a useful fiction, that it is a useless fiction, or that it is the truth. These views will be considered in turn. First we note some recent signs of the growing awareness of the problem.

In the 1984, the subject made its way into a standard text [135]. R. Wald writes: "The breakup of the metric into a background metric which is treated classically and a dynamical field γ_{ab}, which is quantized, is unnatural from the viewpoint of classical general relativity. Furthermore, the perturbation theory one obtains from this approach will, in each order, satisfy causality conditions with respect to the background metric η_{ab} rather than the true metric g_{ab}. Although the summed series (if it were to converge) still could satisfy appropriate causality conditions, the covariant perturbation approach would provide a very awkward way of displaying the role of the spacetime metric in causal structure." [135] (p. 384). Once again, a prior commitment to a geometrical understanding of classical gravity is evident. Some of Wald's negative attitude toward the "breakup" of g_{ab} results from assuming that the curved metric is fundamental, not derived. But given how easy it is to *derive* Einstein's equations from a flat spacetime theory [17, 112], why should one not regard the *curved* metric as derived? Be that as it may, one is pleased that the light cone issue is emerging from the neglect that it once suffered. It is intriguing that Wald suggests that the whole series might be g_{ab}-causal even though each term is η_{ab}-causal. An easy way for such to occur would be for the curved metric's null cone in fact to be confined on or within the flat one's. If that is the case, then it seems that Wald is almost suggesting (albeit reluctantly) what we will do below.

The recent contemplation of "naive quantum gravity" by S. Weinstein also has called attention to the lack of a fixed causal structure in quantum gravity [169]. If one is interested in full quantum gravity, as opposed to semiclassical work, then "we would expect that the metric itself is subject to quantum fluctuations ... But if the metric is [subject to quantum fluctuations], then it is by no means clear that it will be meaningful to talk about whether x and y are spacelike separated, unless the metric fluctuations somehow leave the causal (i.e. conformal) structure alone." (pp. 96-7) There appear to be two things that this last suggestion might mean. First, it might mean that the metric is conformally flat, so that the causal structure is just that of flat spacetime, while gravity is described by a scalar field. However, it is well-known that scalar gravity is empirically falsified by the classical tests of general relativity [17]. Second, it might mean that, although the full metric is allowed to vary, its variations are *bounded* so that the null cone of the nondynamical (and presumably flat) metric is respected. That is what we propose here. Weinstein does indeed consider "whether it is at all *possible* to construe gravitation as a universal interaction that nonetheless propagates in flat, Minkowski spacetime." (p. 91) He concludes that

> the short answer is, 'No,' for three reasons. First, the 'invisibility' of the flat spacetime means that there is no privileged way to decompose a given curved spacetime into a flat background and a curved perturbation about that background. Though this

> non-uniqueness is not particularly problematical for the classical theory, it is quite problematical for the quantum theory, because different ways of decomposing the geometry (and thus retrieving a flat background geometry) yield different quantum theories. Second, not all topologies admit a flat metric, and therefore spacetimes formulated on such topologies do not admit a decomposition into flat metric and curved perturbation. Third, it is not clear *a priori* that, in seeking to make a decomposition into background and perturbations about the background, the background should be *flat*. For example, why not use a background of constant curvature? (p. 92)

However, these arguments seem less than compelling. Concerning the first argument, Weinstein provides neither argument nor citation. It appears to be a claim that a suitably gauge-invariant theory cannot be constructed. Supposing that this claim is true, which is not obvious, it might be cause for "gauge-fixing" the theory at a fundamental level, a proposal which has in fact already been endorsed at the classical level by N. Rosen [3], A. Papapetrou [8], A. A. Logunov and collaborators such as A. A. Vlasov (for example, [106]), and H. Nikolić [109], or perhaps for adding mass term to the theory, if the negative energy and causality worries discussed elsewhere can be handled. But one would like to know the justification for the claim that a gauge-invariant theory is impossible. Concerning the second objection, which resembles that of Ashtekar and Geroch, the advocate of flat spacetime will ask "why are nontrivial topologies necessary?" There are no *facts* or even good arguments that require them at present.[82] In the absence of such, the insistence that nontrivial topologies are theoretically necessary is close to question-begging. For why not merely adopt a nongeometrical view of gravity at the classical level, too?[5] Concerning the last objection, it seems clear that a flat background is the default choice because it is simpler than any other choice. While any other choice requires some argument for making that choice instead of the others and strongly suggests the question "why does spacetime have *this* geometry?", flat spacetime does not, but rather is the obvious default choice. Weinstein's specific alternative suggestion of a constant curvature spacetime, for example, suggests the question "why does the curvature take *this* value, as opposed to some other value?" We can agree with Weinstein that in "allowing metric fluctuations to affect causal structure, one is clearly at some remove from ordinary field-theoretic quantization schemes." (p. 97) But it seems unclear, *pace* Weinstein, that there is any need to renounce the use of a flat background causal structure.

It is unfortunate that some recent articles still do not address the null cone issue. However, enough have done so that one can identify three major attitudes toward the use of a flat metric that one finds. One view is that the flat metric is a useful fiction. Another, more purely geometrical view holds that the flat metric is a useless fiction. The third view regards the flat

[5]The issue of the topology of the universe has seen a fair amount of attention lately, with the aim of experimental test–see, for example, [170–175]. We thank Dr. C. Verjovsky Marcotte for the first two references.

metric as the truth. We survey these approaches in this order.

2.4 Field Formulation: the Flat Metric as a Useful Fiction

Some authors have explicitly stated that the flat metric is merely an auxiliary object, formally useful but not tied to the causal structure of the theory [84–86, 103, 104]. The reasons given include the gauge-variance of the relationship between the null cones and the unobservability of the flat metric. The fact that the flat metric's null cone is sometimes violated using otherwise-convenient gauges appears to be another reason: it does not appear possible to fix the gauge to be, say, tensorial DeDonder and have the null cone relationship be automatically satisfactory. L. P. Grishchuk has written that "the mutual disposition of the light cones of the $g_{\mu\nu}$ and $\eta_{\mu\nu}$ can be of interest only in the case when the attempt is made to interpret the metric relations of the world as observable," which efforts are of course bound to fail, he continues [86]. Unfortunately, Grishchuk has overlooked the same distinction that Gupta, Kraichnan, Thirring, Feynman, and probably many others missed, and has failed to recognize that, if the null cones can be made consistent, then a conceptual difficulty posed by quantization would be eliminated. A. N. Petrov describes the same view (though we have taken the liberty of spelling out with words the abbreviations used):

> However, the background in the field formulation of general relativity is not observed. The movement of test particles and light rays is not connected with the geometry of the background

> spacetime. The light velocity in the background spacetime can approach an infinite value. In contrast, in the geometrical formulation of general relativity the test particles and the light rays define the geodesics in real physical spacetime. Thus, the background spacetime in the field formulation of general relativity is an auxiliary and nonphysical (fictitious) concept which is necessary for the description of true physical fields [104] (pp. 452, 453).

We admit inability to understand how a fictitious entity could be "necessary for the description of true physical fields". It would appear that if the object in question is necessary for the description of true physical fields, then it is real; but if it is not real, then it is not necessary for the description of true physical fields. But let us continue with Petrov:

> We stress that the field formulation of general relativity and the geometrical formulation of general relativity are two different formalisms for a description of the same physical reality and they lead to the same physical conclusions ... there are no obstacles in treating any solution to general relativity (spacetime) in the framework of the field formulation of general relativity. However, it is clear that a manifold which supports a physical metric will not coincide in general with a "manifold" which supports an auxiliary metric. As a result, in the field configuration on the auxiliary nonphysical background, "singularities," "mem-

> branes," "absolute voids," and others can appear. This leads to cumbersome and confused interpretations and explanations. Thus, the whole spirit of general relativity itself requires the investigation of many problems with the help of the geometrical formulation technique. However, there exists problems [sic] for an investigation in which the field formulation technique is more convenient [104] (p. 453).

Thus, the field formulation is seen as a tool that sometimes is helpful, but sometimes not so convenient, and in any case not to be trusted in addressing deep issues. If the flat metric tensor is to be praised so faintly, one might wonder if it is worth using.

A similar attitude has been taken by D. E. Burlankov [87], who did some early work using a flat background metric as a convenient fiction [35]. Burlankov objects to the fundamental status of Minkowski spacetime because of the gauge-variance of the null cone relation, and also because the curved null cone *differs* from the flat null cone [87]. The former argument will be addressed in due time. The latter argument, in Burlankov's hands, is said to imply that only curved metrics conformally related to the flat background would be acceptable. But this objection is just unpersuasive. It is not worrisome if the gravitational field slows light down below the universal velocity constant, as long as gravity does not speed light up. Burlankov's position [87] is fairly similar to that of Zel'dovich and Grishchuk, but a few points deserve special notice. Burlankov is sympathetic to idea (asserted by

Logunov *et al.*) that general relativity has difficulties, noting "the collapse problem, the singularity problem, strong gauge invariance, and the absence of a 'natural' energy-momentum complex" (p. 176). However, Burlankov finds that the "solution of the amazing problems in gravity does not lie" in the bimetric formalism (p. 177): Minkowski space cannot be taken as fundamental because of the null cone difficulties.

2.5 Geometrical Formulation: the Flat Metric as a Useless Fiction

Other authors have taken the view that the flat metric is a blemish on the pure geometric beauty of general relativity, and thus is to be avoided in general. Such a description would seem to fit R. Penrose [81], J. Bičák [162], and L. Shepley[6]. This negative attitude toward the flat metric seems to have motivated Penrose to note that the null cone issue really must be handled if the Lorentz-covariant approach (which he associates with Weinberg) is to be considered satisfactory. Penrose, recognizing the connection between scattering theory and the Lorentz-covariant perturbation approach to gravity, poses a dilemma for the latter. Using global techniques, he shows that either the curved null cone locally violates the flat one, or the scattering properties become inconvenient because the geodesics for the two metrics continue to diverge even far away from a localized source. He concludes that a "satisfactory" relationship between the two null cones cannot be found. Concerning the horns of Penrose's dilemma, we simply accept the second one. It is

[6]We thank Prof. L. Shepley for discussing this issue.

known that long-range fields have inconvenient scattering properties [166]. We find that the root of the divergence between the geodesics is merely the long-range $\frac{1}{r}$ character of the potential in the conformally invariant part of the curved metric. If the fall-off were a power law of the form $\frac{1}{r^{1+\epsilon}}$, $\epsilon > 0$, then no difficulty would arise. So this objection is basically a reflection of the fact that a long-range symmetric tensor potential exists. But why is that a fundamental problem?

2.6 Special Relativistic Approach: the Flat Metric as the Truth

Besides the field-in-fictitious flat spacetime and geometrical approaches, there is another attitude that one might take toward the flat metric approach, *viz.*, that special relativity is correct in its usual strict sense (global Lorentz invariance, trivial spacetime topology, and no violation of η-causality), and thus that the gravitational field must be made to respect the flat metric's causal structure. This view is more conservative than the other views [40] (p. 101), and is sufficiently obvious and attractive an idea that one might expect it to have been explored thoroughly, probably decades ago, and either sorted out or refuted. But as a matter of fact, we do not find that to be the case. Demonstrating this surprising fact was one purpose of the substantial review of the history of the subject above. Some authors have claimed to have sorted it out, and some to have refuted it, but we disagree on both points, as will appear below.

For the sake of convenience, this approach needs a name. We will use the term "special relativistic approach" (SRA). We have resisted calling this approach a "formulation" to match Petrov's "field formulation" and "geometric formulation," because it will appear below that the SRA is in fact physically distinct, though in rather subtle and recondite ways, from the geometrical approach. Some have objected to regarding a theory based on the Einstein equations as something other than general relativity [84–86]. Others have insisted that the SRA is distinct from general relativity[7]. Perhaps the common usage of the term "general relativity" is simply too vague to provide a resolution to this difficulty. If nothing else were at stake, one would avoid pretentious claims of a new theory. But as will appear below, there is in fact be a physical difference, which might even be testable, between the two approaches. In particular, the phenomena of collapse to form black holes seem to be altered somewhat in the SRA, not least because the SRA implies global hyperbolicity.

Some terminology will be helpful. Let us agree to refer to lengths and times measured using ordinary rods and clocks, which respond to $g_{\mu\nu}$, as "physical" lengths and times, but lengths and times measured using ideal rods and clocks, which are not affected by gravity and thus respond to $\eta_{\mu\nu}$, "metaphysical" lengths and times. Let us also write ds^2 for the g-interval and $d\sigma^2$ for the η-interval. The terms "physical" and "metaphysical" are parallel to the more traditional terms "renormalized" and "unrenormalized," but we find our terms more descriptive of our viewpoint. Suppose that one

[7]We thank Prof. Shepley for discussing this issue.

wants to carry a wristwatch around in a gravitational field in order to know what time it is and interact with others. It is not obvious simply from the terms used that one wants a "renormalized" watch, but it is obvious that one wants a "physical" watch in order to live successfully in the physical world. On the other hand, suppose that one wants to know whether to take seriously the infinite character of $t = \infty$ in naive coordinates for the Schwarzschild solution in the SRA. It is not obvious that one would trust a "unrenormalized" clock, and indeed many authors have heaped insults on the readings of such ethereal clocks. But it is indeed obvious that a "metaphysical" clock is to be trusted about ∞ in preference to a merely "physical" clock. Our terminology is thus well-suited to a program of taking the taking the flat metric, including its causal structure, seriously, a goal perhaps not shared by some authors who have trafficked in "renormalized" and "unrenormalized" measurements. It should be clear that while both the field formulation and the special relativistic approach use a flat background metric and share many mathematical results, the SRA takes a realist attitude toward the flat metric, whereas the field formulation takes an antirealist attitude.

The SRA has the advantage of simple and fixed notion of causality at the classical level, because the flat null cone serves as a bound on the curved one. In this view, one becomes less dependent on the study of topology, global techniques, careful definitions of causality of various sorts, and the like, such as modern texts contain [129, 135]. The SRA is therefore simpler in an obvious way than the alternatives. There is no difficulty in extending this nondynamical causal structure into the quantum regime, so there should be no problem in writing down equal-time commutation relations, *etc.* in the usual way. Thus, worries expressed by Ashtekar and Geroch, Isham, van Nieuwenhuizen, Wald and Weinstein above are resolved.

While the special relativistic approach to Einstein's equations is *locally* and *classically* equivalent to the usual theory, as we saw earlier, there might be different *global* or *quantum* properties. For example, though flat spacetime with trivial topology is stable in general relativity [177], closed flat space is unstable [178], so it appears that the usual topology is more than just a simple and convenient choice for the SRA. Also, it will turn out that some regions of spacetime in complete exact solutions of the geometrical theory simply do not exist in the SRA, perhaps due to infinite postponement from the lapse's tending toward 0..[8] Moreover, the SRA of general relativity has less gauge freedom than the geometrical and field formulations, because any gauge choice that leads to an improper null cone relationship must be prohibited. (To be more precise, the SRA configurations form a proper subset of the naive configurations, but with the same order of infinite size.) The use of a new set of variables, in which only the null-cone respecting field configurations are possible, would be a way to prohibit them. We propose such a set below.

The attitude of regarding the flat spacetime as fundamental has been most visibly promoted by A. A. Logunov and colleagues [90–92] (to name

[8] These results are in preparation.

a few).[9] To distinguish their view clearly from any geometrical notions, they have given the name "relativistic theory of gravitation" (RTG) to the work. The nature of the RTG has evolved slightly over the years. For some time it consisted in Rosen's tensorial $\Gamma\Gamma$ action for general relativity and his tensorial DeDonder condition [3] postulated as necessary, presumably with specification of trivial topology for spacetime. There is also attached a "causality principle" that requires that the curved null cone not violate the flat one [90, 164, 165]. This causality principle, which seems to have appeared following criticisms by Zel'dovich and Grishchuk [85, 86], is the feature most relevant to our purposes.[10]

Logunov *et al.*, being committed to the flat spacetime view, regard the question of compatible null cones as requiring a solution. Furthermore, they believe it to be solved already by their causality principle, which we shall call the Logunov Causality Principle. The Logunov Causality Principle states that field configurations that make the curved metric's null cone open wider than the flat metric's are physically meaningless [90, 164, 165]. As they observe, satisfaction is not guaranteed (even with their gauge conditions, notes Grishchuk [86]), which means that the set of partial differential equations is not enough to define the theory. The Logunov Causality Principle is therefore enforced "by hand." Some causality principle is indeed needed, but

[9] This school has also produced an energetic critique of geometrical general relativity as lacking physical meaning, in the sense of lacking conservation laws and failing to make definite predictions. We do not endorse this critique.

[10] More recently, the RTG has sometimes featured acquired a mass term, but we are interested especially in the massless version.

the Logunov Causality Principle strikes us as somewhat arbitrary and *ad hoc*. One would desire three improvements. First, one would prefer that the causality principle be tied somehow to the Lagrangian density or field variables, not separately appended [86]. Second, one wants a guarantee that there exist enough solutions obeying the principle to cover all physically relevant situations. Third, one would prefer a more convenient set of variables to describe the physics. We address all of these matters below.

Concerning the first shortcoming, it might be suggested [92] that the Logunov causality principle is analogous to the energy conditions [135] that one typically imposes. However, this analogy strikes us as weak. The dissimilarity is in how the two conditions accept or reject solutions. The energy conditions are used to exclude or include whole classes of matter fields, so any configuration with one sort of matter field–perhaps a minimally coupled massless scalar field with the correct sign in the Lagrangian density–is permitted, whereas any configuration with another sort of matter field–perhaps the scalar field with the wrong sign– is prohibited as unphysical. This criterion expresses the idea that some sorts of matter are physically reasonable, but others are not. Furthermore, there is no worry that a permissible sort of matter could evolve into a forbidden sort in accord with the field equations. On the other hand, the Logunov Causality Principle cannot give (or at least has not given) a similarly general explanation for why it rejects some solutions of the field equations. A more serious problem is that it cannot give any assurance that it permits a sufficiently large number of solutions to cover all physical situations that arise.

Let us focus our attention on the second difficulty, the possible shortage of solutions, which is potentially very serious. *A priori* there is no reason to believe that one can (partially) fix the gauge, and then still reject some solutions in the appropriate gauge as unphysical. *A posteriori* there seems to be good evidence that this worry is serious. We will employ a somewhat homely example because of its obvious physical relevance. Suppose that a young man named Nicholas has a drum set and a pair of sticks. If Nicholas is a skilled drummer, then the motion of his sticks will be quite under his control, but nevertheless the position of his sticks as a function of time will be rather wild and violent from a kinematical point of view. In particular, we can have great confidence that the motion of his sticks (and arms) will be such that their quadrupole moment will have a nonvanishing second time derivative, in general. We can also be confident that the traceless part of this second time derivative, contracted with itself, tends not to vanish. But that means that Nicholas emits gravitational radiation, for this is just the formula for the average power radiated in general relativity, under suitable assumptions such as a slowly varying source [135]. There will be anisotropic but roughly spherical waves of gravitational radiation diverging from Nicholas. Far away from him, these waves will look approximately like plane waves obeying linearized gravity with the tensorial Hilbert gauge condition (or approach such waves for high frequencies, in the massive case). The behavior of plane monochromatic single-polarization waves in linearized general relativity is well-known [182]. In this gauge, the two energy-carrying polarizations both consist of alternately shrinking one transverse direction and stretching the other, while the time (lapse) and spatial propagation directions are unaffected [182]. (Below we will analyze this linear solution using a generalized eigenvalue-eigenvector formalism.) But the shrinking of one spatial eigenvalue while leaving the time lapse unaltered implies a violation of the flat metric's null cone. In short, it appears that, if the exact behavior of the plane waves is anything like the linearized behavior, then monochromatic gravitational radiation satisfying the tensorial DeDonder condition generically violates the Logunov causality principle. While monochromatic radiation is a rather idealized case, and thus perhaps need not obey η-causality by itself for a satisfactory theory, it is not at all obvious that the superposition of monochromatic waves which all individually violate η-causality yields a sufficiently generous set of realistic waves that satisfy that condition. Thus, there is reason to worry that the Logunov causality principle cannot be implemented in worlds in which Nicholas plays drums.

Evidently, arranging for wave solutions to obey the causality principle is rather more difficult than addressing most of the solutions that Logunov and collaborators have addressed to date, which often have considerable symmetry and fairly trivial dynamics. Regarding standard homogeneous and isotropic cosmological solutions, perhaps one could learn to accept a requirement that the scale factor just could not take certain values [163].

Concerning the Kerr-Newman solution [164], the exterior causes no problem, and the physical significance of the vacuum interior is open to question–especially given a mass term or the reduction of the gauge freedom, as the RTG requires. As far as Kasner solutions are concerned [165], the world looks little like a Kasner solution; maybe somehow it just *couldn't have* looked like a Kasner solution–it's hard to say. Perhaps one could accept these requirements. But no one will accept the idea that Nicholas is unable to play his drums, because he does it every day. It is therefore rather likely that the Logunov causality principle does not admit enough solutions to account for manifestly physically relevant situations, such as Nicholas's drumming. Thus, some way of enforcing null cone consistency without excluding necessary solutions of the field equations must be sought.

The lesson that we draw from the apparent shortcomings of the RTG in its present and past forms is not, *pace* some authors [84–86], that the flat metric must be considered merely a useful fiction. Rather, if the SRA is to be maintained, then a more fundamental approach to securing consistency between the null cones must be sought. One will want to use the gauge freedom of general relativity to secure null cone consistency. In that way, one can be confident that a sufficiently large number of solutions exist, because one member from each equivalence class of solutions will be included.

3 Describing and Enforcing the Proper Null Cone Relationship

3.1 Consistent Null Cones by Suitable Gauge Restrictions?

As several authors above pointed out, the local relation between the two null cones is indeed gauge-dependent in general relativity [81, 86]. One might therefore hope to design a set of gauge-fixing conditions that yield the desired behavior, or at least to impose restrictions on the variables that exclude unsuitable gauge choices while permitting suitable ones. Rather than putting the conditions in arbitrarily by hand, one prefers to implement them in the action principle or the choice of field variables somehow.

It might be hoped that the ADM split of the metric [134, 135], which is quite useful in applications and in identifying the true degrees of freedom, would be a good language for discussing the null cone consistency issue. Let us see if that is the case. For convenience we choose Cartesian coordinates, so that $\eta_{\mu\nu} = diag(-1, 1, 1, 1)$. We therefore make an ADM split of Logunov's 4-dimensional analysis of the causality principle. In considering whether all the vectors V^μ lying on η's null cone are g-timelike, g-null, or g-spacelike, it suffices to consider future-pointing vectors with unit time component; thus $V^\mu = (1, V^i)$, where $V^i V^i = 1$ (the sum running from 1 to 3). The causality principle can be written $h_{ij}(\beta^i + V^i)(\beta^j + V^j) - N^2 \geq 0$ for all spatial unit vectors V^i. Here the spatial metric is h_{ij}, the lapse is N, and the shift is β^i. One could visualize this equation as the requirement that an ellipsoid (not centered at the origin if the shift β^i is nonzero) not protrude from the closed unit ball. Unfortunately, the "all" in "for all spatial unit vectors" is not too easy to handle, so we will in fact look for a better language than an ADM split for discussing null cone consistency.

If there to be any is hope for restricting the gauge freedom so as to ensure that the curved null cone stays consistent with the flat one, then there must be "enough" gauge freedom to transform any physically significant solution into a form that satisfies η-causality. Here we argue that gauges satisfying the causality principle likely *do* exist, so there is enough gauge freedom. Given a flat background metric and a Cartesian coordinate system for it, one can readily draw the flat and curved metrics' light cones on the tangent space at some event (apart from obvious difficulties with higher-dimensional pictures). One wants the curved cone to be located on or within the flat one. The flat cone has the usual ideal conical shape, whereas the curved one is distorted and tilted, in general. In a bimetric context, it is basically the case that the curved spatial metric controls the width of the light cone, while the shift vector determines its tilt from the vertical (future) direction and the lapse function determines its length. (Given that only the conformal part of the metric affects the null cone, the ADM description is a bit redundant.) For general relativity, the spatial metric contains the physical degrees of freedom; the lapse and shift represent the gauge freedom, so they can be chosen arbitrarily, at least over some region. By analogy with conditions typically imposed in geometrical general relativity to avoid causality difficulties [135], one would prefer, if possible, that the curved light cone be strictly inside the flat light cone (*i.e.*, be η-timelike), not tangent to it, because tangency indicates that the field is on the verge of η-causality violation. Under quantization, one might expect fluctuations to push the borderline case into the unacceptable realm, so it seems best to provide a cushion to avoid the problem, if possible. This requirement we call "stable η-causality," by analogy to the usual condition of stable causality [135]. One might worry that this requirement would exclude all curved metrics conformally related to the flat one, and even the presumed "vacuum" $g_{\mu\nu} = \eta_{\mu\nu}$ itself. This worry is justified, but if one takes the message of gauge invariance seriously, then there is no fundamental basis for preferring $g_{\mu\nu} = \eta_{\mu\nu}$ over having the curved metric agree *up to a gauge transformation* with the flat metric. With this relaxed criterion, one avoids the troubles with the folded surface discussed below.

Let the desired relation between the null cones hold at some initial moment. Also let the curved spatial metric and shift be such at some event in that moment that they tend to make the curved cone violate the flat one a bit later. By suitably reducing the lapse, one can lengthen the curved cone until it once again is safely inside the flat cone. By so choosing the lapse at all times and places, one should be able to satisfy the causality principle at every event, if no global difficulties arise. In a rough sense, one might use up $\frac{1}{4}$ of the gauge freedom of general relativity, while leaving the remainder. It is possible that this procedure forces the lapse toward 0 in some cases, which implies that physical events are stretched out over more and more of Minkowski spacetime, perhaps to future or past infinity.

Generally it is assumed that the reason that the gravitational Hilbert action is gauge-invariant is because such gauge invariance reflects a deep feature of the world. However, as we saw above, one can give a somewhat humbler explanation: it is known from the flat spacetime approach that eliminating the time-space components of the field is essential for positive energy properties in Lorentz-invariant theories [1, 75], though in fact the time-time component need not be [80]. We might suggest that the gauging away of the time-time component is necessary rather to respect η-causality. So gauge invariance appears to be required to respect positive energy and special relativistic causality.

3.2 The Causality Principle and Loose Inequalities

As should be clear from the worries about conformally flat curved metrics, the desired relationship between the two null cones takes the form of some loose inequalities $a \leq b$. Such relations have been called "unilateral" [183–187] or "one-sided" [188–190], typical examples being nonpenetration conditions. Such constraints are rather more difficult to handle than the standard "bilateral" or "two-sided" constraint *equations* that most treatments of constraints in physics discuss. Loose inequalities are also more difficult to handle than strict inequalities $a < b$, such as the positivity conditions in canonical general relativity [191, 192, 207–209], which require that the "spatial" metric be spatial. One might eliminate the positivity conditions by a change of variables [192] that satisfies the inequalities identically, such as an exponential function $h = e^y$, as Klotz contemplates. While that is possible, it does not come for free: the "ground state" in which the curved metric equals the flat one is not permitted for any finite value of the argument. However, given the gauge freedom, this objection does not seem compelling, even if it complicates linearization. (Using an even function such as cosh to solve the inequality would seem to introduce even greater difficulties, so we prefer the exponential form.)

If one does leave causality constraints in the theory, rather than solving them as was suggested in the previous paragraph, then one must worry about impulsive constraint forces. That would be the case if one lets the curved metric evolve freely until it "hits" the flat null cone, at which point it might "bounce off," or perhaps become deformed like a face against a window, neither of which seems like an appealing prospect. Another option might be to make the constraint "ineffective", so the constraint force vanishes on account of the constraint itself [197]. However, it seems considerably more satisfactory to reduce the configuration space of the theory so that the problem is avoided altogether, without imposing any gauge-fixing equations at all.

3.3 New Variables and the Segré Classification of the Curved Metric with Respect to the Flat

Previous formulations of the causality principle, which have used the metric [90, 91, 164, 165] or ADM variables as above, have been sufficiently inconvenient to render progress difficult. This was our third complaint about Logunov's formulation of η-causality. One could achieve a slight savings by using the conformally invariant weight $-\frac{1}{2}$ densitized part of the metric $g_{\mu\nu}(-g)^{-\frac{1}{4}}$. Then nine numbers at each event are required (the determinant being -1), which is a bit better than the 10 of the full metric, but still too many.

One would like to diagonalize $g_{\mu\nu}$ and $\eta_{\mu\nu}$ simultaneously by solving the generalized eigenvalue problem

$$g_{\mu\nu}V^{\mu} = \Lambda\eta_{\mu\nu}V^{\mu}, \tag{1}$$

or perhaps the related problem using $g_{\mu\nu}(-g)^{-\frac{1}{4}}$. However, in general that is impossible, because there is not a complete set of eigenvectors, due to the Lorentzian nature of the metric [198–202] (and references in ([202]). There are four Segré types for a real symmetric rank 2 tensor with respect to a Lorentzian metric, the several types having different numbers and sorts of eigenvectors [198–201].

To our knowledge, the only previous work to consider a generalized eigenvector decomposition of a curved Lorentzian metric with respect to a flat one[11] was done by I. Goldman [203], in the context of Rosen's bimetric the-

[11] There is also a literature on choosing coordinates to diagonalize a curved metric [204].

ory of gravity, which does not use Einstein's field equations and is a quite distinct theory. Goldman's work was tied essentially to Rosen's theory, so it does not address our concerns much. The lack of gauge freedom in Rosen's theory also ensured that the curved null cone was not subject to adjustment, unlike the situation in general relativity with a flat metric. As it happens, in Rosen's theory, for static spherically symmetric geometries, the causality principle is always *violated*[12], so that theory does not qualify for a SRA. However, Einstein's theory does evidently have enough gauge freedom to make a special relativistic approach possible.

An eigenvalue-eigenvector decomposition for the *spatial* metric was briefly contemplated by Klotz and Goldberg [191, 192]. For space, as opposed to spacetime, one has a positive definite background (identity) matrix, so the usual theorems apply. But Klotz and Goldberg, who did not assume a flat metric tensor to exist, found little use for the eigenvector decomposition because of the nontensorial nature of the 3×3 identity matrix. Such a decomposition, even given a flat metric tensor, is still somewhat complicated if the ADM shift is nonvanishing ($g_{0i} \neq 0$), as it usually is. Diagonalization has been quite useful in the study of spatially homogeneous cosmologies [196], but our interest is not in specific solutions only, but the general case. However, one expects that studying such models would be instructive.

According to K. P. Tod, "there is very little change for Lorentzian metrics" compared to Riemannian ones [204]. But there is a large change for the eigenvalue problem of interest to us in changing from a Riemannian to a Lorentzian background metric, so the connection between these problems must be somewhat loose.

[12]Prof. Goldman has kindly provided this information from his dissertation in Hebrew.

Let us now proceed with the diagonalization project. Given that a complete set of generalized eigenvectors might fail to exist, it is necessary to consider how many eigenvectors do exist and under which conditions. This problem has been substantially addressed in a different context by G. S. Hall and collaborators [198–201], who were interested in classifying the stress-energy or Ricci tensors with respect to the (curved) metric in (geometrical) general relativity. Such problems have in fact been studied over quite a long period of time [202] (and references therein), but we find the work of Hall *et al.* to be especially convenient for our purposes. There exist four cases, corresponding to the four possible Segré types (apart from degeneracies) for the classified tensor. The case $\{1,111\}$ has a complete set of eigenvectors (1 timelike, 3 spacelike with respect to η), and is thus the most convenient case. The case $\{211\}$ has two spacelike eigenvectors and one null eigenvector (with respect to η), whereas the $\{31\}$ case has one spacelike eigenvector and one null one. The last case, $\{z\,\bar{z}11\}$ has 2 (real) spacelike eigenvectors and 2 complex eigenvectors.

We now consider the conditions under which metrics of each of these Segré classes obey η-causality. To give a preview of our results, we state that the $\{1,111\}$ and $\{211\}$ cases sometimes do obey it, although the $\{211\}$ metrics appear to be dispensable. But no metric of type $\{31\}$ or $\{z\,\bar{z}11\}$ obeys the causality principle, so these types can be excluded from consideration for the SRA.

Hall *et al.* introduce a real null tetrad of vectors $L^\mu, N^\mu, X^\mu, Y^\mu$ with vanishing inner products, apart from the relations $\eta_{\mu\nu}L^\mu N^\nu = \eta_{\mu\nu}X^\mu X^\nu = \eta_{\mu\nu}Y^\mu Y^\nu = 1$, so L^μ and N^μ are null, while X^μ and Y^μ are spacelike. (The signature is $- + ++$.) Expanding an arbitrary vector V^μ as $V^\mu = V^L L^\mu + V^N N^\mu + V^X X^\mu + V^Y Y^\mu$ and taking the η-inner product with each vector of the null tetrad reveals that $V^L = \eta_{\mu\nu}V^\mu N^\nu$, $V^N = \eta_{\mu\nu}V^\mu L^\nu$, $V^X = \eta_{\mu\nu}V^\mu X^\nu$, and $V^Y = \eta_{\mu\nu}V^\mu Y^\nu$. Thus, the Kronecker delta tensor can be written as $\delta^\mu_\nu = L^\mu N_\nu + L_\nu N^\mu + X^\mu X_\nu + Y^\mu Y_\nu$, indices being lowered here using $\eta_{\mu\nu}$. For some purposes it is also convenient to define the timelike vector $T^\mu = \frac{L^\mu - N^\mu}{\sqrt{2}}$ and the spacelike vector $Z^\mu = \frac{L^\mu + N^\mu}{\sqrt{2}}$.

We employ the results of Hall *et al.* [198–201], who find that the four possible Segré types (ignoring degeneracies) for a (real) symmetric rank 2 tensor in a four-dimensional spacetime with a Lorentzian metric can be written in the following ways, using a well-chosen null tetrad. The type $\{1,111\}$ can be written as

$$g_{\mu\nu} = 2\rho_0 L_{(\mu}N_{\nu)} + \rho_1(L_\mu L_\nu + N_\mu N_\nu) + \rho_2 X_\mu X_\nu + \rho_3 Y_\mu Y_\nu, \tag{2}$$

or equivalently

$$g_{\mu\nu} = -(\rho_0 - \rho_1)T_\mu T_\nu + (\rho_0 + \rho_1)Z_\mu Z_\nu + \rho_2 X_\mu X_\nu + \rho_3 Y_\mu Y_\nu. \tag{3}$$

As usual, the parentheses around indices mean that the symmetric part should be taken [135]. The type $\{211\}$ can be written as

$$g_{\mu\nu} = 2\rho_1 L_{(\mu}N_{\nu)} + \lambda L_\mu L_\nu + \rho_2 X_\mu X_\nu + \rho_3 Y_\mu Y_\nu, \tag{4}$$

with $\lambda \neq 0$, the null eigenvector being L^μ. The type $\{31\}$ can be written as

$$g_{\mu\nu} = 2\rho_1 L_{(\mu}N_{\nu)} + 2L_{(\mu}X_{\nu)} + \rho_1 X_\mu X_\nu + \rho_2 Y_\mu Y_\nu, \tag{5}$$

the null eigenvector again being L^μ. The final type, $\{z\,\bar{z}11\}$, can be written as

$$g_{\mu\nu} = 2\rho_0 L_{(\mu}N_{\nu)} + \rho_1(L_\mu L_\nu - N_\mu N_\nu) + \rho_2 X_\mu X_\nu + \rho_3 Y_\mu Y_\nu, \qquad (6)$$

with $\rho_1 \neq 0$. The requirements to be imposed upon the curved metric for the moment are the following: all η-null vectors must be g-null or g-spacelike, all η-spacelike eigenvectors must be g-spacelike, $g_{\mu\nu}$ must be Lorentzian (which amounts to having a negative determinant), and $g_{\mu\nu}$ must be connected to $\eta_{\mu\nu}$ by a succession of small changes which respect η-causality and the Lorentzian signature. It convenient to employ a slightly redundant form that admits all four types in order to treat them simultaneously. Thus, we write

$$g_{\mu\nu} = 2AL_{(\mu}N_{\nu)} + BL_\mu L_\nu + CN_\mu N_\nu + DX_\mu X_\nu + EY_\mu Y_\nu + 2FL_{(\mu}X_{\nu)}. \quad (7)$$

Using this form for $g_{\mu\nu}$, one readily finds the squared length of a vector V^μ to be

$$g_{\mu\nu}V^\mu V^\nu = 2AV^L V^N + B(V^N)^2 + C(V^L)^2 + D(V^X)^2 + E(V^Y)^2 + 2FV^X V^N. \qquad (8)$$

It is not clear *a priori* how to express sufficient conditions for the causality principle in a convenient way. Obviously it is sufficient that every g-timelike vector be η-timelike and every g-null vector be η-null or η-timelike. One could alternatively say that every η-spacelike vector must be g-spacelike and every η-null vector be g-null or g-spacelike. However, timelike and spacelike vectors are not very convenient to use because of the inequalities inherent in the word "every". But it will turn out that in four dimensions, the necessary conditions that we can readily impose are also sufficient.

3.4 Necessary Conditions for Respecting the Flat Metric's Null Cone

The causality principle requires that the η-null vectors L^μ and N^μ be g-null or g-spacelike, so $B \geq 0, C \geq 0$. These conditions already exclude the type $\{z\,\bar{z}11\}$, because the form above requires that B and C differ in sign. It must also be the case that the η-spacelike vectors X^μ and Y^μ are g-spacelike, so $D > 0$ and $E > 0$.

Not merely L^μ and N^μ, but all η-null vectors must be g-null or g-spacelike. This requirement quickly implies that $E \geq A$, and also requires that

$$B(V^N)^2 + 2FV^XV^N + (D - A)(V^X)^2 \geq 0. \tag{9}$$

Here there are two cases to consider: $F = 1$ for type $\{31\}$, and $F = 0$ for types $\{1,111\}$ and $\{211\}$. Let us consider $F = 1$. The $\{31\}$ has $B = 0$, so the equation reduces to $2FV^XV^N + (D - A)(V^X)^2 \geq 0$, which implies that either $V^X = 0$ or, failing that, $2FV^N + (D - A)V^X \geq 0$. Clearly one could also consider a null vector with the opposite value of V^X, yielding the inequality $2FV^N - (D - A)V^X \geq 0$. Adding these two inequalities gives $4V^N \geq 0$, which simply cannot be made to hold for all values of V^N. Thus, the $F = 1$ case yields no η-causality-obeying curved metrics, and the $\{31\}$

type is eliminated. It remains to consider $F = 0$ for the $\{1, 111\}$ and $\{211\}$ types. The resulting inequality is $B(V^N)^2 + (D - A)(V^X)^2 \geq 0$. Because $B \geq 0$ has already been imposed, it follows only that $D \geq A$.

Let us summarize the results so far. The inequalities $B \geq 0$ and $C \geq 0$ have excluded the type $\{z\,\bar{z}11\}$. We also have $D > 0$, $D \geq A$, $E > 0$, $E \geq A$. Finally, $F = 0$ excludes the type $\{31\}$, so only $\{1, 111\}$ and $\{211\}$ remain.

We now impose the requirement of Lorentzian signature. At a given event, there can be no objection to finding a coordinate x such that $(\frac{\partial}{\partial x})^\mu = x^\mu$ and (flipping the sign of y^μ if needed for the orientation) a coordinate y such that $(\frac{\partial}{\partial y})^\mu = y^\mu$; these two coordinates can be regarded as Cartesian. Then the null vectors L^μ and N^μ lie in the $t-z$ plane of this sort of Cartesian system. The curved metric has a block diagonal part in the $x-y$ plane with positive determinant, so imposing a Lorentzian signature means ensuring a negative determinant for the 2×2 $t-z$ part. The vectors L^μ and N^μ in one of these coordinate systems take the form $L^\mu = (L^0, 0, 0, L^3)$ and $N^\mu = (N^0, 0, 0, N^3)$. Given the Cartesian form $\eta_{\mu\nu} = diag(-1, 1, 1, 1)$ and the nullity of these two vectors, it follows that $|L^0| = |L^3|$ and $|N^0| = |N^3|$. Therefore the relevant parts of the curved metric can be written in such a coordinate basis as

$$\begin{aligned} g_{00} &= 2AL^0N^0 + B(L^0)^2 + C(N^0)^2, \\ g_{03} = g_{30} &= -A(N^0L^3 + L^0N^3) - BL^0L^3 - CN^0N^3, \\ g_{33} &= 2AL^3N^3 + B(L^3)^2 + C(N^3)^2. \end{aligned} \quad (10)$$

Taking the determinant using *Mathematica* and recalling that $|L^0| = |L^3|$

and $|N^0| = |N^3|$, one finds that the condition for a negative determinant is $2(A^2 - BC)|L^3|^2|N^3|^2(sign(L^0L^3N^0N^3) - 1) < 0$. The linear independence of L^μ and N^μ implies that $sign(L^0L^3N^0N^3) = -1$, so the determinant condition is $A^2 - BC > 0$. Because B and C are both nonnegative, $A^2 - BC > 0$ implies that $A \neq 0$. But the requirement that the curved metric be smoothly deformable through a sequence of signature-preserving steps means that the curved metric's value of A cannot "jump" from one sign of A to another, but must agree with the flat metric's positive sign. It follows that $A > 0$.

We now summarize the necessary conditions imposed:

$$A > 0, \quad A^2 > BC, \quad B \geq 0,$$
$$C \geq 0, \quad D \geq A, \quad E \geq A,$$
$$F = 0. \tag{11}$$

3.5 Sufficient Conditions for Respecting the Flat Metric's Null Cone

Thus far, we have shown nothing of the sufficiency of these necessary conditions. We now prove that these conditions are in fact sufficient. It is helpful to consider the two types, $\{1, 111\}$ and $\{211\}$, separately.

For the type $\{1, 111\}$, the conditions on the coefficients A, B, *etc.* reduce to

$$A > 0, \quad A > B, \quad B \geq 0,$$
$$C = B, \quad D \geq A, \quad E \geq A. \tag{12}$$

For this form the following relations between variables hold:

$$A = \rho_0, \quad B = \rho_1, \quad D = \rho_2,$$
$$E = \rho_3. \tag{13}$$

It follows that this type can be expressed as

$$g_{\mu\nu} = -(A - B)T_\mu T_\nu + (A + B)Z_\mu Z_\nu + DX_\mu X_\nu + EY_\mu Y_\nu. \tag{14}$$

Writing the eigenvalues for T^μ, X^μ, Y^μ, and Z^μ as D^0_0, D^1_1, D^2_2, and D^3_3, respectively, one has

$$D^0_0 = A - B, \quad D^1_1 = A + B, \quad D^2_2 = D,$$
$$D^3_3 = E. \tag{15}$$

One sees that the inequalities imply that the eigenvalue for the timelike eigenvector T^μ (briefly, the "timelike eigenvalue") is no larger than any of the spacelike eigenvalues:

$$D^0_0 \le D^1_1, \quad D^0_0 \le D^2_2, \quad D^0_0 \le D^3_3, \tag{16}$$

and that all the (generalized) eigenvalues are positive. Let us now see that these conditions are sufficient. Writing an arbitrary vector V^μ as $V^\mu = V^T T^\mu + V^X X^\mu + V^Y Y^\mu + V^Z Z^\mu$, one sees that its η-length (squared) is $\eta_{\mu\nu} V^\mu V^\nu = -(V^T)^2 + (V^X)^2 + (V^Y)^2 + (V^Z)^2$. Clearly this length is never more positive than $\frac{1}{D^0_0} g_{\mu\nu} V^\mu V^\nu = -(V^T)^2 + \frac{D^1_1}{D^0_0}(V^X)^2 + \frac{D^2_2}{D^0_0}(V^Y)^2 + \frac{D^3_3}{D^0_0}(V^Z)^2$, so the necessary conditions are indeed sufficient for type $\{1, 111\}$.

For the type $\{211\}$, the conditions on the coefficients A, B, *etc.* reduce to

$$\begin{aligned} A > 0, \quad B > 0, \quad C = 0, \\ D \geq A, \quad E \geq A, \quad F = 0. \end{aligned} \tag{17}$$

One can write the curved metric in terms of T^μ, Z^μ, X^μ, and Y^μ, though T^μ and Z^μ are not eigenvectors. One then has

$$g_{\mu\nu} = -(A - \frac{1}{2}B)T_\mu T_\nu + (A + \frac{1}{2}B)Z_\mu Z_\nu + BZ_{(\mu}T_{\nu)} + DX_\mu X_\nu + EY_\mu Y_\nu. \tag{18}$$

Writing an arbitrary η-spacelike vector field V^μ as $V^\mu = GT^\mu + HZ^\mu + IX^\mu + JY^\mu$, with $H^2 + I^2 + J^2 > G^2$, one readily finds the form of $g_{\mu\nu}V^\mu V^\nu$. Employing the relevant inequalities and shuffling coefficients, one obtains the manifestly positive result $g_{\mu\nu}V^\mu V^\nu = A(H^2 + I^2 + J^2 - G^2) + \frac{1}{2}B(G - H)^2 + (D - A)I^2 + (E - A)J^2$. This positivity result says that all η-spacelike vectors are g-spacelike. Earlier the requirement that all η-null vectors be g-null or g-spacelike was imposed. These two conditions together comprise the causality principle, so we have obtained sufficient conditions for the $\{211\}$ type also.

The $\{211\}$ type, which has with one null and two spacelike eigenvectors, is a borderline case in which the curved metric's null cone is tangent to the flat metric's cone along a single direction [205]. Clearly such borderline cases of $\{211\}$ metrics obeying the causality principle form in some sense a measure 0 set of all causality principle-satisfying metrics. Given that they are so scarce, one might consider neglecting them. Furthermore, they are arbitrarily close to violating the causality principle. We recall the criterion of stable causality in geometrical general relativity [135] (where the issue is closed timelike curves, without regard to any flat metric's null cone), which frowns upon metrics which satisfy causality, but would fail to do so if perturbed by an arbitrarily small amount. One could imagine that quantum fluctuations might push such a marginal metric over the edge, and thus prefers to exclude such metrics as unphysical. By analogy, one might impose stable η-causality, which excludes curved metrics that are arbitrarily close to violating the flat null cone's notion of causality, though we saw that such a condition would exclude conformally flat metrics, also. Perhaps a better reason for neglecting type $\{211\}$ metrics is that they are both technically inconvenient and physically unnecessary. Because η-causality-respecting $\{211\}$ metrics are arbitrarily close to $\{1,111\}$ metrics, one could merely make a gauge transformation[13] to shrink the lapse a bit more and obtain a $\{1,111\}$ metric instead. Thus, every curved metric that respects η-causality either is of type $\{1,111\}$, or is arbitrarily close to being of type $\{1,111\}$ and deformable thereto by a small gauge transformation reducing the lapse.

It follows that there is no loss of generality in restricting the configuration space to type $\{1,111\}$ curved metrics, for which the two metrics are simultaneously diagonalizable. As a result, there exists a close relationship

[13] Here we refer to gauge transformations in the field formulation, where the flat metric's null cone is ignored. It will become evident that the notion of a gauge transformation in the special relativistic approach, which respects η-causality, is more restrictive.

between the traditional orthonormal tetrad formalism and this eigenvector decomposition. In particular, one can build a g-orthonormal tetrad field e^{μ}_{A} simply by choosing the normalization of the eigenvectors. This choice removes the local Lorentz freedom of the tetrad (except when eigenvalues are degenerate).

Rewriting the generalized eigenvector equation for the case in which a complete set exists, one can write $g_{\mu\nu}e^{\mu}_{A} = \eta_{\mu\nu}e^{\mu}_{B}D^{B}_{A}$, with the four eigenvalues being the elements of the diagonal matrix D^{A}_{B}. It is sometimes convenient to raise or lower the indices of this matrix using the matrix $\eta_{AB} = diag(-1, 1, 1, 1)$. The tetrad field has $\{e^{\mu}_{A}\}$ has inverse $\{f^{A}_{\mu}\}$. We recall the standard relations $g_{\mu\nu} = f^{A}_{\mu}\eta_{AB}f^{B}_{\nu}$ and $g_{\mu\nu}e^{\mu}_{A}e^{\nu}_{B} = \eta_{AB}$. It is not difficult to show the how the tetrad lengths are related to the eigenvalues: $\eta_{\mu\nu}e^{\mu}_{A}e^{\nu}_{B} = D^{-1}_{AB}$, and equivalently, $\eta^{\mu\nu}f^{A}_{\mu}f^{B}_{\nu} = D^{AB}$. It follows that $f^{A}_{\nu} = \eta_{\nu\alpha}e^{\alpha}_{B}D^{AB}$, which says that a given leg of the cotetrad f^{A}_{ν} can be expressed solely in terms of the corresponding leg of the tetrad e^{μ}_{A}, through a stretching, an index lowering, and possibly a sign change, without reference to the other legs. Simultaneous diagonalization implies that the tetrad vectors are orthogonal to each other with respect to *both* metrics.

It would be interesting to explore the use of these eigenvectors as the orthonormal tetrad field in C. Møller's tetrad formalism. Concerning localization of gravitational energy, Møller concluded that a satisfactory solution within Riemannian geometry does not exist, but that one does exist in a tetrad form of general relativity, apart from the question of finding six extra equations to fix the freedom under local Lorentz transformations, because the localization does not depend on the world coordinate choice or the global Lorentz frame [143–148]. In the case of the present eigenvector formalism, the additional six equations require the several eigenvectors to be η-orthogonal to each other. It will appear below that spatial eigenvalue degeneracy would render 3 of these equations less useful than one might wish, but in any case the remaining 3 can always be imposed.

3.6 Linearized Plane Waves a Difficulty for the Logunov Causality Principle

As was stated in connection with Nicholas's drumming, monochromatic plane gravitational waves satisfying the linearization of the Einstein equations and the Hilbert (linearized DeDonder) gauge appear to violate η-causality in general. We will now show that in more detail, using Ohanian and Ruffini [182] as our guide, while making use of the eigenvalue technology introduced above. These results will also hold approximately for the Maheshwari-Logunov massive theory for large frequencies and weak fields.

Defining a trace-reversed potential $\phi^{\mu\nu} = \gamma^{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}\gamma$ (with $\gamma^{\mu}_{\mu} = \gamma$) and imposing the Hilbert gauge $\partial_{\mu}\phi^{\mu\nu} = 0$, one puts the linearized Einstein equations in the form $\partial^2\phi^{\mu\nu} = 0$. Because we desire plane wave solutions, let $\lambda\phi^{\mu\nu} = H\epsilon^{\mu\nu}\cos(k_{\alpha}x^{\alpha} + \psi)$, where $\epsilon^{\mu\nu}$ is a constant polarization tensor, k^{α} a constant polarization vector, and H is a small number fixing the amplitude. We let the waves travel in the z-direction, so $k^{\mu} = \omega(1, 0, 0, 1)$. We also define the vectors $\epsilon_1^\mu = (0,1,0,0)$ and $\epsilon_2^\mu = (0,0,1,0)$.

The gauge condition implies that the polarization tensor is orthogonal to the propagation vector: $\epsilon^{\mu\nu} k_\mu = 0$, leaving six independent solutions. One can take the six independent polarization tensors (with both indices lowered using the flat metric) to be:

$$\epsilon_{1\mu\nu} = \epsilon_{1\mu}\epsilon_{1\nu} - \epsilon_{2\mu}\epsilon_{2\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{19}$$

$$\epsilon_{2\mu\nu} = \epsilon_{1\mu}\epsilon_{2\nu} + \epsilon_{2\mu}\epsilon_{1\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{20}$$

$$\epsilon_{3\mu\nu} = \epsilon_{1\mu}\frac{1}{\omega}k_\nu + \epsilon_{1\nu}\frac{1}{\omega}k_\mu = \begin{bmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{21}$$

$$\epsilon_{4\mu\nu} = \epsilon_{2\mu}\frac{1}{\omega}k_\nu + \epsilon_{2\nu}\frac{1}{\omega}k_\mu = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{22}$$

$$\epsilon_{5\mu\nu} = k_\mu k_\nu = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \tag{23}$$

$$\epsilon_{6\mu\nu} = \epsilon_{1\mu}\epsilon_{1\nu} + \epsilon_{2\mu}\epsilon_{2\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{24}$$

The first two, which are transverse-transverse and traceless, are the physical (energy-carrying) polarizations in the massless theory. One can show that they induce effective curved metrics of type $\{1111\}$, but, partly due to the oscillations of the cosine function, they violate η-causality, as the behavior of the eigenvalues shows. Let us be rather explicit and find an η-null vector that is on occasion g-timelike for the first polarization. The polarization tensor being traceless, one has

$$g_{\mu\nu} = \eta_{\mu\nu} - \lambda\gamma_{\mu\nu} = diag(-1, 1, 1, 1) - H\cos(k_\alpha x^\alpha + \psi) diag(0, 1, -1, 0). \tag{25}$$

The vector $U^\mu = (1, 1, 0, 0)$ is η-null, but is g-timelike during every other half-period of the cosine function. Thus, this physical polarization violates η-causality.

Perhaps at this stage one could hope that the gauge waves will be more friendly toward η-causality, and that by including enough of the gauge waves, one could tame the bad properties of the physical waves. The polarizations $\epsilon_{3\mu\nu}$ and $\epsilon_{4\mu\nu}$, which are transverse-longitudinal and traceless, however, will probably not help, because the resulting metrics are of type $\{31\}$, and thus violate η-causality. The fifth polarization $k_\mu k_\nu$, which is longitudinal-longitudinal and traceless, is of type $\{211\}$, and thus could satisfy η-causality, but the fluctuating cosine implies that if causality is respected during one half-period, then it is violated during the next. So it does not help much either. The last polarization $\epsilon_{1\mu}\epsilon_{1\nu} + \epsilon_{2\mu}\epsilon_{2\nu}$ is transverse-transverse but not traceless. This wave is of type $\{1111\}$, but the oscillating eigenvalues once again violate η-causality.

Thus, all six polarizations of plane wave admitted by the linearized Einstein equations in the Hilbert (linearized DeDonder) gauge individually violate η-causality. It would be surprising if it is possible to make sufficiently general (for Nicholas's drumming, for example) superpositions of these waves that respect η-causality, given that each one violates it individually. Let us consider this question more explicitly. One can consider a general superposition

$$\lambda\gamma_{\mu\nu} = \sum_{A=1}^{6} H_A \epsilon_{A\mu\nu} \cos(k_\alpha x^\alpha + \psi_A) - \eta_{\mu\nu} H_6 \cos(k_\alpha x^\alpha + \psi_6), \qquad (26)$$

where ψ_A are real phases. The characteristic polynomial $|g_{\mu\nu} - \Lambda\eta_{\mu\nu}| = 0$ can be simplified by defining $B_A = H_A \cos(k_\alpha x^\alpha + \psi_A)$ (with no summation over A). The result is

$$\begin{vmatrix} 1+B_5+B_6-\Lambda & -B_3 & -B_4 & -B_5 \\ -B_3 & -1+B_1+\Lambda & B_2 & B_3 \\ -B_4 & B_2 & -1-B_1+\Lambda & B_4 \\ -B_5 & B_3 & B_4 & -1+B_5-B_6+\Lambda \end{vmatrix} = 0. \tag{27}$$

It looks difficult to get practical results from this equation in the general case, it being a messy quartic polynomial. However, the third and fourth polarizations, affecting the only shift and an off-diagonal spatial component each, look rather unlikely to help satisfy η-causality, but probably would harm it. It seems fair to set $B_3 = B_4 = 0$. The characteristic polynomial then becomes block diagonal, with roots $\Lambda = 1 \pm \sqrt{B_1^2 + B_2^2}$ and (repeated) $\Lambda = 1 + B_6$. If $B_5 \neq 0$, then $\Lambda = 1 + B_6$ corresponds to the null eigenvector $(1, 0, 0, 1)$, yielding a $\{211\}$ metric (which at best only just satisfies causality), but setting $B_5 = 0$ gives another eigenvector and hence a $\{1111\}$ metric, which can satisfy stable causality, so we choose the latter. However, it is clear that one of the other two eigenvalues will at almost every moment be less than unity if a physical wave is present, while $1 + B_6$ will be greater than unity half the time. Thus, plausibly, a general superposition of gauge and physical polarizations will not rescue η-causality.

Clearly this argument is not a mathematical proof, and does not even use exact solutions of Einstein's equations. However, this argument does make it seem likely that plane wave-like solutions, which are necessary near the future light cone of Nicholas's drum set, violate η-causality. If that conclusion is correct, then the RTG's approach to causality indeed must be modified, as we urged above.

3.7 Dynamics of the Causality Principle

It is one thing to know the kinematic inequalities to describe the causality principle, but another to enforce them. A difficulty comes from trying to mate the causality principle with the condition that, as the gravitational field gets very weak,

$$g_{\mu\nu} \rightarrow \eta_{\mu\nu}, \tag{28}$$

which seems like a natural requirement on a flat spacetime theory of gravity. (However, it is not gauge-invariant.) Given that the derivation of general relativity involves a defining relation of the form $g_{\mu\nu} = \eta_{\mu\nu} - \lambda\gamma_{\mu\nu}$ [112] or the like for the curved metric in terms of the flat and the gravitational potential $\gamma_{\mu\nu}$, clearly $g_{\mu\nu} = \eta_{\mu\nu}$ is expected to be the state of no gravity (apart from gauge transformations).

The causality principle inequalities can be written as

$$D^0_0 \leq D^1_1, \quad D^0_0 \leq D^2_2, \quad D^0_0 \leq D^3_3. \tag{29}$$

One can abbreviate this set of inequalities by

$$D^0_0 \leq \min(D^1_1, D^2_2, D^3_3). \tag{30}$$

In three spacetime dimensions one can readily plot the function $D^0_0 \leq \min(D^1_1, D^2_2)$ using *Mathematica*. The result is seen in two spatial dimensions in figure 2.

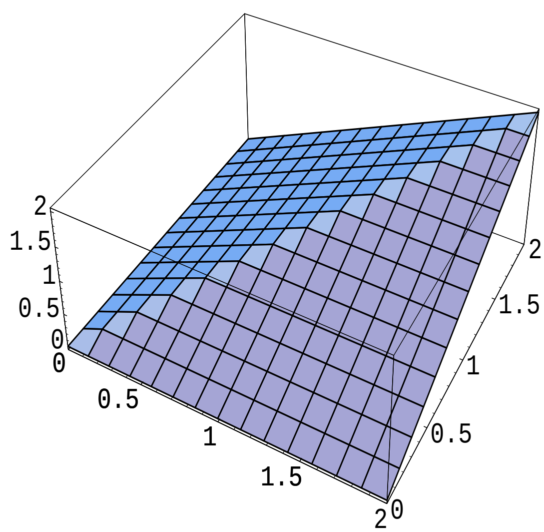

Figure 2: Bounding Surface for Temporal Eigenvalue as Function of Spatial Eigenvalues in 2 Dimensions

The greatest allowed values of the timelike eigenvalue lie on the folded surface shown. In three spatial dimensions, a similar (but more complicated) situation arises.

Perhaps motivated by the suggestion that bimetric theories ought to be gauge-fixed, one might attempt to enforce the causality principle by (partial) gauge-fixing, perhaps setting the timelike eigenvalue equal to some function of the spacelike eigenvalues only. This would be an algebraic or "ultralocal" gauge fixing. One might want to choose a specific function $D^0_0 = f(D^1_1, D^2_2, D^3_3)$ to fix the gauge. If $f(D^1_1, D^2_2, D^3_3) = \min(D^1_1, D^2_2, D^3_3)$, then the curved null cone is always on the verge of violating the flat cone, but never quite does so. By analogy with the requirement of stable causality in the standard causal analysis of geometrical general relativity, one would prefer that the metric not be arbitrarily close to unacceptable behavior, because a quantization might lead to fluctuations that push it over the edge [135]. A more attractive choice is therefore

$$f(D^1_1, D^2_2, D^3_3) = \frac{\min^2(D^1_1, D^2_2, D^3_3)}{\max(D^1_1, D^2_2, D^3_3)}, \tag{31}$$

which provides a cushion except when the curved metric equals the flat one. This choice also corresponds to the usual form of the Schwarzschild solution, while being consistent with the explicitly conformally flat form of the flat Robertson-Walker cosmological model [135].

If one were to write these functions explicitly, they would contain Heaviside step functions. Thus they are merely continuous, which is a rather weak level of differentiability. This difficulty is in fact inevitable. If the selected gauge at all permits the D^0_0 to take its maximum value along the fold, then the gauge can be continuous, but clearly is not twice continuously differentiable, as one would wish. Because the lapse is closely related to D^0_0, a merely continuous timelike eigenvalue implies a merely continuous lapse. The lapse being gauge freedom, one might hope that continuity is enough for the lapse. Unfortunately, the Einstein equations (in Hamiltonian form) contain second spatial derivatives of the lapse [135]. Two derivatives of a merely continuous function will yield a Dirac delta function in the $\dot{\pi}^{ab}$ equation. So the momentum will evolve discontinuously if D^0_0 ever touches the fold. Such behavior seems physically unreasonable, so the fold must be avoided if a sensible result is to emerge. Avoiding the fold under all circumstances means that a flat spacetime theory of gravity does not have $g_{\mu\nu} = \eta_{\mu\nu}$ as a solution, which seems surprising at first, but the gauge-variance of $g_{\mu\nu} = \eta_{\mu\nu}$ implies that this loss is not too worrisome.

In any case, if one wishes to fix the gauge, there is no need to restrict attention to such *ultralocal* forms as $D^0_0 = f(D^1_1, D^2_2, D^3_3)$, with no derivatives present. Why not a merely local gauge fixing, with a finite number of derivatives? Analogously, in geometrical general relativity, one considers not only Gaussian normal coordinates $g^{00} = -1, g^{0i} = 0$, but also harmonic coordinates, and in electromagnetism, one considers not only the temporal gauge $A_0 = 0$ [206], but also the Lorentz gauge. However, even given this enlarged set of gauge-fixing choices, it is not clear that there exists any way to include the solution ($g_{\mu\nu} = \eta_{\mu\nu}$) without encountering discontinuous evolution, which seems unacceptable.

It seems best to avoid these difficulties altogether at the theoretical level by abstaining from *fixing* the gauge (and instead merely *restrict* the gauge to respect η-causality), and to take the gauge freedom seriously enough to give up the solution $g_{\mu\nu} = \eta_{\mu\nu}$ as inessential. One might even aim to satisfy the causality-related restrictions on the gauge *identically* using a new set of variables adapted to this purpose.

3.8 Satisfying η-causality Identically Using New Variables

If one imposes stable η-causality, which excludes all metrics that satisfy η-causality but which are arbitrarily close to violating it, along with all that violate it, then the loose inequalities of the causality principle are changed to strict inequalities. These inequalities are then susceptible to being *solved and eliminated*, as Klotz proposed to do with the positivity conditions of canonical gravity [191, 192]. Let us see how this goal can be achieved. For generality, let us work in d spacetime dimensions, where $d = 2$, $d = 3$, and $d = 4$ are perhaps of the most interest, unless one is considering Kaluza-Klein theory. (For $d \neq 4$, one might want to verify that nothing disturbing happens regarding the eigenvector formalism happens, although that seems unlikely.) Given stable η-causality, there always exists a complete set of generalized eigenvectors with real eigenvalues and eigenvectors, as we proved earlier. Some time ago J. A. Schouten cryptically wrote, “The theorems of principal axes and of principal blades do not hold if the fundamental tensor is indefinite because in this case a real symmetric tensor or bivector may possibly have a special position with respect to the real nullcone. But if the symmetric tensor or bivector does not have any such special position the theorems remain valid." [176] (p. 47) Evidently stable η-causality prevents $g_{\mu\nu}$ from having a "special position" with respect to the fundamental tensor $\eta_{\mu\nu}$.

It is clear that the determinant of the effective curved metric, as long as it remains negative, is irrelevant to the relation between the two null cones. One could then split the curved metric into the (conformally invariant) unimodular part $g_{\mu\nu}/\sqrt[d]{-g}$ and the determinant g. However, it is no worse in Cartesian coordinates, and perhaps more convenient in non-Cartesian coordinates, to use the *covariantly* unimodular metric $\hat{g}_{\mu\nu} = g_{\mu\nu}/\sqrt[d]{\frac{g}{\eta}}$ and the coordinate scalar $\kappa = \sqrt{\frac{g}{\eta}}$.

Let us consider the generalized eigenvector problem in terms of the covariantly unimodular metric:

$$\hat{g}_{\mu\nu} u^{\mu} = \eta_{\mu\nu} u^{\mu} \hat{\Lambda}. \tag{32}$$

Given that we have already required that the curved metric have a complete set of eigenvectors with respect to the flat, we can multiply the relation $\hat{g}_{\mu\nu} u^{\mu}_{A} = \eta_{\mu\nu} u^{\mu}_{B} \hat{D}^{B}_{A}$ (where the diagonal matrix $\hat{D}^{B}_{A}$ has unit determinant, rendering $\hat{D}^{0}_{0}$ dependent on the other components) by the inverse matrix U^{-1} of the matrix U of unnormalized eigenvectors u^{μ}_{A}. The result is

$$\hat{g}_{\rho\nu} = \eta_{\mu\nu} u^{\mu}_{B} \hat{D}^{B}_{A} u^{-1A}_{\rho}. \tag{33}$$

If this equation is twice contracted with the eigenvalue matrix U, giving

$$\hat{g}(u_A, u_C) = \eta(u_B, u_C)\hat{D}^B_A, \tag{34}$$

then the freedom to normalize the eigenvectors can be put to work. We have already that the eigenvectors to be orthogonal (with respect to both metrics). It is useful to make u^μ_A be η-unit vectors.[14]

Let us now specialize to Cartesian coordinates and write u^C_A, if necessary, to remind ourselves of that fact. The normalization relation $\eta_{\mu\nu} u^\mu_A u^\nu_B = \eta_{AB}$, if expressed in Cartesian coordinates, is nothing other than the condition [213] for U to be in the complete Lorentz group $O(1,3)$ if $d = 4$. It is clear on physical grounds that the gravitational field ought not to reverse time or space, so only the subgroup $SO(1,3) \uparrow$, the proper orthochronous Lorentz group with $|U| = 1$ and $u^0_0 \geq 0$, is relevant. This subgroup being connected to the identity, any of its matrices u^C_A can be obtained by exponentiating a matrix in the Lie algebra of the Lorentz group, that is, the algebra of real matrices W^A_B such that $\eta_{CA} W^A_B$ is antisymmetric [213]. Given this η-unit normalization convention and the use of Cartesian coordinates, we can use the relations

$$\begin{aligned} U &= e^W, \\ U^{-1} &= e^{-W} \end{aligned} \tag{35}$$

to eliminate the eigenvectors in favor of a matrix W describing how the

[14] Earlier we saw that g-unit normalization was convenient for tying the eigenvectors to the standard orthonormal tetrad formalism. These two different normalizations are adapted to different purposes.

eigenvectors are boosted or rotated relative to the coordinate basis. The orthogonality of the eigenvectors is satisfied identically using this construction. If one wishes, it is possible to write W in terms of a standard representation of the infinitesimal generators of the Lorentz group with known commutation relations [214] (pp. 538-541). Setting $d \neq 4$ would of course require using the Lorentz group in the appropriate dimensions. Recently progress with this sort of expression has been made [210].

Let us consider the eigenvalues $\hat{D}^A_B$. We saw above that the timelike eigenvalue for the covariantly unimodular metric is dependent upon the spacelike eigenvalues. : $\hat{D}^0_0 = \frac{1}{\prod_{i=1}^{d-1} D^i_i}$. One can define reduced eigenvalues

$$\bar{D}^A_B = \frac{\hat{D}^A_B}{\hat{D}^0_0} = \frac{D^A_B}{D^0_0}; \tag{36}$$

clearly $\bar{D}^0_0 = 1$. It is interesting that now only three numbers (compared to nine components of $\hat{g}_{\mu\nu}$) are needed to describe the null cone relationship sufficiently to show whether η-causality is respected. Let us now eliminate $\hat{D}^A_B$ in favor of the reduced eigenvalues using the relation $\hat{D}^0_0 = \frac{1}{\sqrt[d]{|\bar{D}|}}$. At this stage, one can write the (covariantly) unimodular metric in matrix form as

$$\hat{g} = \frac{e^{-W} \bar{D} e^{W} \eta}{\sqrt[d]{|\bar{D}|}}. \tag{37}$$

While one could represent W by a 2-form [210] if one wished, which would permit the use of an arbitrary coordinate system, this eigenvalue formalism requires the introduction of a set of Cartesian coordinates for the eigenvalue matrix.

Imposing stable η-causality gives the *strict* inequalities

$$\bar{D}^1_1 > 1, \bar{D}^2_2 > 1, \bar{D}^3_3 > 1. \tag{38}$$

But strict inequalities can be solved, as Klotz suggested [191, 192], and then the causality constraints would be eliminated. This goal is achieved by setting, for example,

$$\bar{D}^1_1 = e^{\alpha} + 1, \bar{D}^2_2 = e^{\beta} + 1, \bar{D}^3_3 = e^{\gamma} + 1, \tag{39}$$

for the case $d = 4$, with obvious changes for other dimensions. Formally defining a matrix $M^B_A = diag(-\infty, \alpha, \beta, \gamma)$ lets one write $\bar{D} = I + e^M$. The covariantly unimodular metric is now $\hat{g} = \frac{(I+e^{-W}e^M e^W)\eta}{\sqrt[d]{|I+e^M|}}$. One can also set $\kappa^{\frac{2}{d}} = e^{\delta}$. The full metric can then be written as

$$g = \frac{e^{\delta}(I + e^{-W}e^M e^W)\eta}{\sqrt[d]{|I + e^M|}}. \tag{40}$$

We let the lower index indicate the row and the upper, the column.

If $d = 2$, it is not difficult to write out this form of the curved metric explicitly. (In higher dimensions, the noncommutativity of the various rotations and boosts in W presents more difficulty, although some simplification has been achieved [210].) The matrix W can be put in the form

$$W = \begin{bmatrix} 0 & \epsilon \\ \epsilon & 0 \end{bmatrix}. \tag{41}$$

One quickly finds that

$$U = \begin{bmatrix} \cosh\epsilon & \sinh\epsilon \\ \sinh\epsilon & \cosh\epsilon \end{bmatrix},$$

$$U^{-1} = \begin{bmatrix} \cosh\epsilon & -\sinh\epsilon \\ -\sinh\epsilon & \cosh\epsilon \end{bmatrix}. \tag{42}$$

Finishing the elementary algebra gives

$$g = \frac{e^{\delta}}{\sqrt{1+e^{\alpha}}} \begin{bmatrix} -1 + e^{\alpha}\sinh^2\epsilon & -e^{\alpha}\cosh\epsilon\sinh\epsilon \\ -e^{\alpha}\cosh\epsilon\sinh\epsilon & 1 + e^{\alpha}\cosh^2\epsilon \end{bmatrix}. \tag{43}$$

With all three (or ten or $\frac{d(d+1)}{2}$) components of the curved metric expressed in a form along these lines, the following highly desirable properties of the curved metric are satisfied *automatically*:

1. the curved metric satisfies stable η-causality;

2. the curved metric has Lorentzian signature;

3. the gravitational field has not reversed time or space;

4. the curved metric respects global hyperbolicity.

It would seem that these are all the properties that one would desire from an effective curved metric in the SRA. The matrix W (apart from the near-antisymmetry) and α, β, γ, and δ can be any real numbers. Thus, as long as a curved metric can be written in terms of these variables, then it is satisfactory for the SRA, and *vice versa.* These variables are evidently well-suited for the SRA applied to a metric theory of gravity. However, it is clear that a technical complication arises when degeneracy among the spatial eigenvalues exists.

If it were not for the frightful amount of calculation involved, and probably the inconveniently long and complicated results of the calculations, it might be interesting to rewrite the entirety of gravitation in the SRA in these variables. One example of the difficulty is what becomes of the primary constraints in the canonical form of the theory, which show that not all 10 components of the metric have canonical momenta of the usual sort suited to Legendre transformations from the Lagrangian density [193]. Until Dirac [194] and J. L. Anderson [195] showed in 1958 how to reduce the primary constraints to the vanishing of some canonical momenta, the presence of primary constraints was a significant issue in canonical general relativity. However, the use of η-causality-adapted variables perhaps prevents reducing the primary constraints to the vanishing of some momenta, so this issue might be reopened.

The issue can be illustrated in a tractable form using a model theory suited to Kasner solutions [191, 192], in which the ansatz

$$g_{\mu\nu}(t, x^i) = diag(g_0(t), g_1(t), g_2(t), g_3(t)) \tag{44}$$

is made before varying the action. The resulting equations in fact reproduce the Einstein equations. In this model theory, the primary constraint is that π^0, the momentum conjugate to g_0, vanish. In the η-causality variables, this diagonal form of the metric is realized by setting $W = 0$ and letting the remaining fields depend on t only. One then obtains for (g_0, g_1, g_2, g_3),

$$\left(\frac{-e^{\delta}}{\sqrt[4]{(1+e^{\alpha})(1+e^{\beta})(1+e^{\gamma})}}, \right.$$
$$\frac{e^{\delta}(1+e^{\alpha})}{\sqrt[4]{(1+e^{\alpha})(1+e^{\beta})(1+e^{\gamma})}},$$
$$\frac{e^{\delta}(1+e^{\beta})}{\sqrt[4]{(1+e^{\alpha})(1+e^{\beta})(1+e^{\gamma})}},$$

$$\left. \frac{e^{\delta}(1+e^{\gamma})}{\sqrt[4]{(1+e^{\alpha})(1+e^{\beta})(1+e^{\gamma})}} \right). \tag{45}$$

Given how the causality variables mix together the coordinates (g_0, g_1, g_2, g_3), one expects the momenta to be comparably mixed, which means that the primary constraint will also mix momenta together.

This model theory might be a good test bed for ascertaining what *explicitly* is the gauge freedom of the SRA. In other words, which vector fields ξ^{μ} generate gauge transformations that respect stable η-causality? With the Lagrangian density rewritten in terms of these η-causality-adapted variables, only this set of vector fields would be admitted as gauge transformations, while any others would be regarded as merely mathematical transformations. One fact that is already clear is that the set of admissible vector fields depends upon the field configuration prior to the transformation, because a vector field that preserves η-causality given one initial field configuration might violate it given another. This situation is rather different from the usual situation (in the geometrical or field formulation), in which just any (reasonable) vector field generates a gauge transformation.

3.9 Finite Gauge Transformations and Orthonormal Tetrads

The form of a *finite* gauge transformation for the densitized inverse metric tensor in the field formulation is known from the work of Grishchuk, Petrov, and A. D. Popova [83] to have the form

$$\mathfrak{g}^{\sigma\rho} \rightarrow e^{\pounds_{\xi}} \mathfrak{g}^{\sigma\rho}, u \rightarrow e^{\pounds_{\xi}} u, \eta_{\mu\nu} \rightarrow \eta_{\mu\nu} \tag{46}$$

in terms of the convenient variable $\mathfrak{g}^{\sigma\rho} = \sqrt{-g}g^{\sigma\rho}$, the flat metric tensor, and matter fields u described by some tensor densities (with indices suppressed).

We recall the bimetric form of the action above for a generally covariant theory, with the metric here expressed in terms of the weight 1 inverse metric:

$$S = S_1[\mathfrak{g}^{\mu\nu}, u] + \frac{1}{2}\int d^4x R_{\mu\nu\rho\sigma}(\eta)\mathcal{M}^{\mu\nu\rho\sigma} + 2b\int d^4x\sqrt{-\eta} + \int d^4x \partial_\mu \alpha^\mu. \quad (47)$$

Clearly the terms other than S_1 either do not change under the assumed field transformation, or do so at most by a boundary term, so our attention turns to $S_1 = \int d^4x \mathcal{L}_1$.

We now derive a useful formula. Writing out $e^{\pounds_\xi}A$ as a series $e^{\pounds_\xi}A = \sum_{i=0}^{\infty} \frac{1}{i!}\pounds_\xi^i A$ for some tensor density A will put us in a position to derive a useful 'product' rule for the exponential of Lie differentiation. One could write a similar series for another tensor density B. Multiplying the series and using the Cauchy product formula [212]

$$\sum_{i=0}^{\infty} a_i z^i \sum_{j=0}^{\infty} b_j z^j = \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_k b_{n-k} z^k \quad (48)$$

and the n-fold iterated Leibniz rule [212]

$$[fg]^{(n)} = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} f^{(k)} g^{(n-k)}, \quad (49)$$

one recognizes the result as the series expansion of $e^{\pounds_\xi}(AB)$, so one has the result

$$(e^{\pounds_\xi}A)(e^{\pounds_\xi}B) = e^{\pounds_\xi}(AB) \quad (50)$$

In view of the matrix relationships among the various metric quantities, one has by definition that $(\mathfrak{g}^{\mu\nu} + \delta\mathfrak{g}^{\mu\nu})(\mathfrak{g}_{\rho\nu} + \delta\mathfrak{g}_{\rho\nu}) = \delta^\mu_\rho$ and various other relations. In that way, one can derive the form of $\delta\mathfrak{g}_{\rho\nu}$, δg, $\delta g_{\rho\nu}$, and the like. Let us show this fact explicitly for g, using $\mathfrak{g}^{\sigma\rho} + \delta\mathfrak{g}^{\sigma\rho} = e^{\pounds_\xi}\mathfrak{g}^{\sigma\rho}$. One knows from mathematics that a determinant is given by $|\mathfrak{g}^{\sigma\rho}| = \epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi\mathfrak{g}^{\alpha\beta}\mathfrak{g}^{\mu\chi}\mathfrak{g}^{\nu\psi}\mathfrak{g}^{\rho\phi}$, where $\epsilon(\alpha\mu\nu\rho)$ is the totally antisymmetric symbol with $\epsilon(0123) = 1$. Because this form for the determinant holds in any coordinate system, $\epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi$ is a *scalar* (and also a constant), so $e^{\pounds_\xi}(\epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi) = \epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi$. We therefore have

$$\begin{aligned}|e^{\pounds_\xi}\mathfrak{g}^{\sigma\rho}| &= \epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi(e^{\pounds_\xi}\mathfrak{g}^{\alpha\beta})(e^{\pounds_\xi}\mathfrak{g}^{\mu\chi})(e^{\pounds_\xi}\mathfrak{g}^{\nu\psi})(e^{\pounds_\xi}\mathfrak{g}^{\rho\phi}) \\ &= e^{\pounds_\xi}|\mathfrak{g}^{\sigma\rho}|. \end{aligned} \tag{51}$$

Using $\mathfrak{g}^{\sigma\rho} = g^{\sigma\rho}\sqrt{-g}$, one finds that $|\mathfrak{g}^{\sigma\rho}| = |g_{\sigma\rho}|$, so

$$g + \delta g = e^{\pounds_\xi} g. \tag{52}$$

The relation $-g - \delta g = (\sqrt{-g} + \delta\sqrt{-g})^2$ defines $\delta\sqrt{-g}$, so one quickly also finds that

$$\sqrt{-g} + \delta\sqrt{-g} = e^{\pounds_\xi}\sqrt{-g}, \tag{53}$$

with which one readily finds the result for $g^{\sigma\rho}$ and so on. Again the transformed field is just the exponentiated Lie derivative of the original.

Grishchuk, Petrov, and Popova have exhibited a straightforward relationship between finite gauge transformations (with the exponentiated Lie differentiation) and the tensor transformation law [100, 101]. Evidently the former is fundamental, the latter derived. One can define a vector field ξ^α using the fact that under a gauge transformation, $\mathfrak{g}^{\mu\nu}$ changes in accord with the tensor transformation law, while the flat metric stays fixed. Let us follow them and define ξ^α in terms of the finite coordinate transformation

$$x'^\alpha = e^{\xi^\mu \frac{\partial}{\partial x^\mu}} x^\alpha. \tag{54}$$

Then the tensor transformation law, which is fairly easy to use, gives the finite gauge transformation formula, which is difficult to use. One can therefore hope to avoid using the latter at all, and work only with the tensor transformation law, which will tend to be simpler and will relate to known results in the geometrical theory.

In case the form for a finite gauge transformation for a tetrad field has not appeared previously, we provide one here. If one imposes no requirements on the tetrad other than that it be orthonormal, then the formula is nonunique in the local Lorentz transformation matrix.[15] It is not difficult to verify that the following form preserves both the completeness relation to the inverse metric $g^{\mu\nu} = e^\mu_A \eta^{AB} e^\nu_B$ and the orthonormality relation $g_{\mu\nu} e^\mu_A e^\nu_B = \eta_{AB}$:

$$e^\mu_A + \delta e^\mu_A = e^{\pounds_\xi} (e^F)^C_A e^\mu_C. \tag{55}$$

The Lie differentiation in the first factor acts on everything to its right. F is a matrix field which, when an index is moved using η_{AB} or η^{AB}, is antisymmetric: $F^C_A = -\eta_{AE} F^E_B \eta^{BC}$. This relation, it should be noted, is not

[15] Two possible ways of adding extra conditions might be: (1) require that the tetrad consist of generalized eigenvectors, or (2) require that the gauge transformation formula follow from an infinite succession of alternate small local Lorentz and gauge transformations.

tied to the presence of a flat background metric tensor or the interpretation of e^{μ}_{A} as generalized eigenvectors.

Let us now verify the completeness relation $g^{\mu\nu} = e^{\mu}_{A}\eta^{AB}e^{\nu}_{B}$, by showing that this relation with the gauge-transformed tetrad yields the gauge-transformed curved metric. The equations $\pounds_{\xi}\eta^{AB} = 0$ and the $(e^{\pounds_{\xi}}A)(e^{\pounds_{\xi}}B) = e^{\pounds_{\xi}}(AB)$ will be used in simplification. One has by definition of a variation Δ induced by this tetrad transformation,

$$\begin{aligned} g^{\mu\nu} + \Delta g^{\mu\nu} &= (e^{\mu}_{A} + \delta e^{\mu}_{A})\eta^{AB}(e^{\nu}_{B} + \delta e^{\nu}_{B}) \\ &= [e^{\pounds_{\xi}}(e^{F})^{C}_{A}e^{\mu}_{C}]\eta^{AB}e^{\pounds_{\xi}}(e^{F})^{E}_{B}e^{\nu}_{E} \\ &= e^{\pounds_{\xi}}[(e^{F})^{C}_{A}e^{\mu}_{C}\eta^{AB}(e^{F})^{E}_{B}e^{\nu}_{E}]. \end{aligned} \tag{56}$$

Acting with $e^{-\pounds_{\xi}}$ gives

$$e^{-\pounds_{\xi}}(g^{\mu\nu} + \Delta g^{\mu\nu}) = (e^{F})^{C}_{A}e^{\mu}_{C}\eta^{AB}(e^{F})^{E}_{B}e^{\nu}_{E}. \tag{57}$$

We shall use the near-antisymmetry of F: $F^{C}_{E} = -\eta_{EJ}F^{J}_{B}\eta^{BC}$. One then has

$$\begin{aligned} e^{-\pounds_{\xi}}(g^{\mu\nu} + \Delta g^{\mu\nu}) &= e^{\mu}_{C}(e^{F})^{C}_{A}\eta^{AB}(e^{F})^{E}_{B}e^{\nu}_{E} \\ &= e^{\mu}_{C}(e^{F})^{C}_{A}\eta^{AB}(I^{E}_{B} + F^{E}_{B} + F^{E}_{J}F^{J}_{B} + \ldots)e^{\nu}_{E}, \end{aligned} \tag{58}$$

where the one factor is expanded as a series. Continuing by moving the Lorentz metric into strategic locations gives

$$\begin{aligned} & e^{\mu}_{C}(e^{F})^{C}_{A}(I^{A}_{P} + \eta^{AB}F^{E}_{B}\eta_{EP} + \eta^{AB}F^{J}_{B}\eta_{JK}\eta^{KL}F^{E}_{L}\eta_{EP} + \ldots)e^{P\nu} \\ &\qquad = e^{\mu}_{C}(e^{F})^{C}_{A}(I^{A}_{J} - F^{A}_{J} + F^{A}_{K}F^{K}_{J} - \ldots)e^{J\nu}, \end{aligned} \tag{59}$$

where the near-antisymmetry of F has been employed. Reverting to the exponential form gives

$$
\begin{aligned}
& e^{\mu}_{C}(e^{F})^{C}_{A}(e^{-F})^{A}_{J}e^{J\nu} \\
& \quad = e^{\mu}_{C} I^{C}_{E} e^{E\nu} \\
& \qquad = g^{\mu\nu},
\end{aligned} \tag{60}
$$

leading to the expected conclusion $g^{\mu\nu} + \Delta g^{\mu\nu} = g^{\mu\nu} + \delta g^{\mu\nu} = e^{\pounds_{\xi}} g^{\mu\nu}$. Thus, completeness holds, and the tetrad-induced variation Δ of the inverse curved metric agrees with the gauge transformation variation δ. By similar maneuvers, one establishes the orthonormality relation for this tetrad variation:

$$
(e^{\pounds_{\xi}} g_{\mu\nu})(e^{\mu}_{A} + \delta e^{\mu}_{A})(e^{\nu}_{B} + \delta e^{\nu}_{B}) = \eta^{AB}. \tag{61}
$$

Finally, the inverse tetrad transforms as

$$
f^{A}_{\mu} + \delta f^{A}_{\mu} = e^{\pounds_{\xi}}(e^{-F})^{A}_{C} f^{C}_{\mu}, \tag{62}
$$

which looks much like the tetrad form, save for the sign of F.

In these relations, we have retained the full local Lorentz freedom. But in the eigenvierbein formalism in the SRA, the vectors, in addition to being g-orthonormal, will be η-orthogonal: $\eta_{\mu\nu} e^{\mu}_{A} e^{\nu}_{B} = 0 if A \neq B$. This set of six orthogonality equations will typically render the tetrad unique, and thus determine the local Lorentz transformation matrix F for a gauge transformation ξ^{μ} (although solving for F does not look easy). However, in view of the difficulties that arise when there is degeneracy among the spatial eigenvalues, it might be preferable to retain manifest rotational invariance for the spatial vectors. Stable η-causality ensures that the timelike eigenvector is always unique, so one is free to require that the timelike leg of the tetrad be just the suitably normalized (and future-pointing) timelike eigenvector of the curved metric with respect to the flat metric. It might be interesting to impose this eigenvector on the timelike leg of the tetrad identically. In that case, the gravitational variables could be taken as the three spatial legs of the cotetrad and the conformal factor $\sqrt{-g}$, 13 numbers at each event. One anticipates that the Dirac formalism would then yield 7 primary constraints for the canonical momenta, along with the usual 4 secondary constraints. One further expects that all 11 the constraints would be first-class, leaving the expected 2 degrees of freedom at each point in space.

3.10 Gauge Transformations Not a Group

If one is not interested in taking η-causality seriously, then any suitably smooth vector field, perhaps subject to some boundary conditions, will generate a gauge transformation. However, in the SRA, respecting η-causality–indeed, preferably stable η-causality–is essential. This fact entails that only a subset of all vector fields generates gauge transformations in the SRA. Let us be more precise in defining gauge transformations in the SRA, requiring stable η-causality. A gauge transformation in the SRA is a mathematical transformation generated by a vector field in the form described above, but which also has both the untransformed and transformed curved metrics respect stable η-causality. It is evident that a vector field that generates a gauge transformation given one curved metric and a flat metric, might not generate a gauge transformation given another curved metric (and the same flat metric), because in the second case, the transformation might move the curved metric out of the stable η-causality-respecting configuration space, which is only a subset of the naive configuration space. It follows that one cannot identify gauge transformations with generating vector fields alone; rather, one must also specify the field configuration (curved metric) assumed prior to the transformation. For thoroughness, one can also use the flat metric (which is not transformed) as a label, to ensure that the trivial coordinate freedom is not confused with the physically significant gauge freedom. Let us therefore provisionally write a gauge transformation as an ordered triple

$$(\xi^{\mu}(x), \eta_{\mu\nu}(x), g_{\mu\nu}(x)), \tag{63}$$

where both $g_{\mu\nu}$(x) and $e^{\pounds_{\xi}} g_{\mu\nu}(x)$ satisfy stable causality with respect to $\eta_{\mu\nu}(x)$. The former restriction limits the configuration space for the curved metric, whereas the latter restricts the vector field. (At this point we drop the indices and the spacetime position argument for brevity.)

One wants to compose two gauge transformations to get a third gauge transformation. At this point, the fact that a gauge transformation is not labelled merely by the vector field, but also by the curved metric (and the flat), has important consequences. Clearly the two gauge transformations to be composed must have the second one start with the curved metric with which the first one stops. We also want the flat metrics to be compatible. Thus, the 'group' multiplication operation is defined only in certain cases, meaning the gauge transformations in the SRA *do not form a group*, despite the inheritance of the mathematical form of exponentiating the Lie differentiation operator from the field formulation's gauge transformation group. Two gauge transformations (ψ, η_2, g_2) and (ξ, η_1, g_1) can be composed to give a new gauge transformation $(\psi, \eta_2, g_2) \circ (\xi, \eta_1, g_1)$ only if $g_2 = e^{\pounds_\xi} g_1$ and $\eta_2 = \eta_1$. Because the multiplication operation is not always defined between two gauge transformations, the closure property of groups fails to hold.

The failure of closure implies modifications of associativity. Associativity does not hold in the usual way, because composition is not always meaningful. However, whenever the composition of SRA gauge transformations is meaningful, associativity holds, due to inheritance from the field formulation gauge group.

The failure of closure also modifies the existence of the identity, at least given the definition of a gauge transformation as an ordered triple. Whereas a group has a single element that acts as the identity on all elements and from either side, gauge SRA transformations do not have any single element that can be multiplied with all other elements, so in particular there is no identity element. There are, rather, many "little identity elements", which all have vanishing vector field, but which differ in their curved (or flat) metric labels. However, this complication can be removed if one alters the definition of a gauge transformation in the following way: let a gauge transformation be an ordered triple of the sort described above if and only if its vector field is nonvanishing, but let there also be a trivial transformation that maps any gauge transformation to itself, never mind any curved or flat metric labels. We therefore identify all transformations $(0, \eta_A, g_B)$ and write them as (0).

The inverse property is essentially untouched by the failure of closure. The left inverse of (ξ, η_1, g_1) is $(-\xi, \eta_1, e^{\pounds_\xi} g_1)$, yielding $(-\xi, \eta_1, e^{\pounds_\xi} g_1) \circ (\xi, \eta_1, g_1) = (0, \eta_1, g_1)$, an identity transformation. The right inverse is also $(-\xi, \eta_1, e^{\pounds_\xi} g_1)$, yielding $(\xi, \eta_1, g_1) \circ (-\xi, \eta_1, e^{\pounds_\xi} g_1) = (0, \eta_1, e^{\pounds_\xi} g_1)$, which is also an identity transformation, or rather, *the* identity transformation (0), given our modified definition of the identity. Thus, every SRA gauge transformation has a two-sided inverse.

In summary, of the group properties of closure, associativity, the existence of an identity element for all elements, and the existence of a two-sided inverse for each element [206], SRA gauge transformations, with the carefully chosen identity transformation, have an identity element for all elements, a two-sided inverse for each element, and associativity in those cases where multiplication is defined, but not closure, because multiplication is not defined for every ordered pair of elements. It turns out that securing the gauge invariance of physical results can imply boundary conditions on the vector field ξ^μ.

3.11 Canonical Quantization in the SRA

The relevance of the special relativistic approach to Einstein's equations to canonical quantum gravity deserves some consideration. The primary patrons of the flat metric in the context of Einstein's equations have been the particle physicists in the context of the old covariant perturbation program of quantization. However, this program famously proved to be nonrenormalizable, even (probably) with the addition of carefully chosen matter fields in the later supergravity era [216]. Therefore, the covariant perturbation program has largely been abandoned.[16]

With the patrons of the flat metric having largely diverted their attention to strings, membranes, and the like, one might form the belief that the use of a flat background metric has nothing further to contribute to quantum gravity, and in particular, to canonical quantum gravity. Isham writes of the null cone issue in the covariant perturbation program: "This very nontrivial problem is one of the reasons why the canonical approach to quantum gravity has been so popular." [216] (p. 12) And again, "One of the main aspirations of the canonical approach to quantum gravity has always been to build a formalism with no background spatial, or spacetime, metric." [216] (p. 18) The use of a flat background in canonical gravity indeed seems to be rather rare, apart from some work in the field formulation by Grishchuk and Petrov [101], which does not consider the flat metric's null cone. (We do not see any reason that the *formal* use of a flat background in the field

[16] An exception is some recent work by G. Scharf and collaborators such as I. Schorn, N. Grillo, and M. Wellmann (for example, [114, 117]). Their use of "causal" methods helps to achieve finite results. Another possibility, suggested some time ago by Weinberg, is that 4-dimensionally nonperturbatively renormalizable. Recently O. Lauscher and M. Reuter have argued that this situation likely is realized [211].

formulation requires one to give up the spatial metric components as the canonical coordinates, as they do.)

However, it would be a mistake to conclude that the canonical formalism is immune to similar worries, worries which a flat background's null cone structure could address. Isham continues:

> However, a causal problem arises here [in the canonical approach] too. For example, in the Wheeler-DeWitt approach, the configuration variable of the system is the Riemannian metric $q_{ab}(x)$ on a three-manifold Σ, and the canonical commutation relations invariably include the set
>
> $$[\hat{q}_{ab}(x), \hat{q}_{cd}(x')] = 0 \tag{64}$$
>
> for all points x and x' in Σ. In normal canonical quantum field theory such a relation arises because Σ is a space-like subset of spacetime, and hence the fields at x and $x\prime$ should be simultaneously measurable. But how can such a relation be justified in a theory that has no fixed causal structure? The problem is rarely mentioned but it means that, in this respect, the canonical approach to quantum gravity is no better than the covariant one. It is another aspect of the 'problem of time' .... [216] (p. 12)

Evidently introducing a flat metric can help:

> The background metric η provides a fixed causal structure with the usual family of Lorentzian inertial frames. Thus, at this level,

> there is no problem of time. The causal structure also allows a notion of microcausality, thereby permitting a conventional type of relativistic quantum field theory ... It is clear that many of the *prima facie* issues discussed earlier are resolved in an approach of this type by virtue of its heavy use of background structure. [216] (p. 17)

What then is the difficulty?

> However, many classical relativists object violently ..., not least because the background causal structure cannot generally be identified with the physical one. Also, one is restricted to a specific background topology, and so a scheme of this type is not well adapted for addressing many of the most interesting questions in quantum gravity: black hole phenomena, quantum cosmology, phase changes *etc.* [216] (p. 17)

However, above we have presented a formalism which, if adopted, plausibly *does* ensure that the physical causal structure is consistent with the background one by construction. Thus, this first objection is largely answered. The second objection is strong only if one already knows that gravitation is geometrical at the classical level. But there is no necessity in taking such a view. We conclude that it would be interesting to investigate the canonical quantization of Einstein's equations within the special relativistic approach, because serious conceptual problems with standard approaches would evidently be resolved, whereas no serious problems would be generated, at least at the conceptual level.

On the other hand, the special relativistic approach might perhaps complicate certain technical issues in canonical quantum gravity. First, as we saw above, the use of the η-causality variables suggests that the primary constraints of the theory would be non-trivialized. Second, the fact that some vector fields do not generate gauge transformations in the SRA suggests that the custom of splitting spatial and temporal diffeomorphisms and treating them independently might be threatened. In the field formulation, with merely formal use of the flat metric, one could at least make sense of spatial gauge transformations as those generated by solutions ξ^{α} of the equation $x'^{\alpha} = e^{\xi^{\mu}\frac{\partial}{\partial x^{\mu}}} x^{\alpha}$ given the restriction that $x'^{0} = x^{0}$. But it is not clear that such a separation is even possible if the flat metric's null cone structure is to be respected, as in the SRA. The reason is that whether a vector field generates a gauge transformation in the SRA depends on *both* the temporal and spatial parts of the vector field, because both influence the curved null cone.

4 Global Hyperbolicity, Black Hole Information Loss, and a Well Posed Initial Value Formulation

With the requirement of η-causality imposed–perhaps stable η-causality using the causality variables–it follows that any "SRA spacetime" $(R^4, \eta_{\mu\nu}, g_{\mu\nu})$

is globally hyperbolic in the sense of Wald [135]. How so? It follows from η-causality that the future domain of dependence of an η-spacelike slice is in fact the whole of R^4. But global hyperbolicity just is the possession of a Cauchy surface [135], so any η-causal SRA spacetime $(R^4, \eta_{\mu\nu}, g_{\mu\nu})$ is globally hyperbolic. But global hyperbolicity resolves the Hawking black hole information loss paradox [180]. That is because "the spacetime can be foliated by a family of Cauchy surfaces in such a way that there is no time at which the Lemma [proved therein] forces the state of the universe to be mixed" [180] (p. 204). Thus, there is no evolution from a pure state to a mixed state, and so no information is lost. If global hyperbolicity solves the problem, then so does the SRA.

If the nature of black holes is altered in the SRA is altered, one might wonder what becomes of the work on black hole entropy. As J. Oppenheim has shown recently [181], the proportionality of black hole entropy to area does not depend on the existence of an event horizon, but merely occurs in the limit as a gravitating system approaches its gravitational radius. Inclusion of the gravitational field in thermodynamics yields a correction term that violates entropy extensivity; in the limit as the radius approaches the Schwarzschild radius, the entropy is proportional to area rather than volume.

One could further ask whether the SRA has a well posed initial value formulation. The use of harmonic coordinates has been a common technique for answering this question in the geometrical formulation [135], in which the choice of harmonic coordinates constitutes a gauge-fixing. Given that the SRA has restricted gauge freedom, and that the tensorial DeDonder gauge condition (which makes the coordinate g-harmonic when η-Cartesian) has causal difficulties for plane wave solutions, one might fear that the proofs of a well posed initial value formulation fail for the SRA. However, such a fear is groundless, as we see if we resist the temptation to use η-Cartesian coordinates, for which we have no need. The SRA has both coordinate freedom and gauge freedom. The choice of g-harmonic coordinates, if nothing is said about the functional form of the flat metric, is merely a coordinate choice, leaving the gauge unfixed. The eigenvalues, which express the relation between the two null cones, are coordinate scalars, and so are indifferent to the choice of g-harmonic coordinates (and possibly messy form for the flat metric). The gauge freedom has not been used at all, and thus is fully available for deforming the curved metric until it becomes consistent with the flat one by reducing the lapse. Therefore the traditional harmonic coordinate approach to demonstrating a well posed initial value problem experiences no obstacles from the SRA. The SRA indeed has a well posed initial value formulation, and one need not even appeal to recent work that permits coordinate freedom [215] to show so.

5 Conclusion

We have aimed to take special relativity seriously while viewing gravity as described by Einstein's field equations. In reviewing the history of the

Lorentz-covariant approach, we found that the fundamental issue of causality has frequently been ignored, and never (so far as we can find) been handled correctly. Next we found an adequate kinematical description of the relation between the null cones, and found that all metrics obeying stable η-causality, and almost all metrics obeying η-causality, possess in effect an orthonormal tetrad of eigenvectors. Using a bit of the gauge freedom of the theory, plausibly one can deforming any physically relevant solution into one in which the proper null cone relation obtains. Having done so, one can adopt a new set of variables which ensure that the proper relation holds automatically. As a result of using the flat metric, the problem of defining causality in quantum gravity is solved. Furthermore, every SRA solution is globally hyperbolic, so the Hawking black hole information loss paradox seems not to arise.

References

[1] M. Fierz, *Helv. Phys. Acta* **12** (1939), 3.

[2] M. Fierz, W. Pauli, *Proc. Roy. Soc. A* **173** (1939), 211.

[3] N. Rosen, *Phys. Rev.* **57** (1940), 147, 150.

[4] N. Rosen, *Ann. Phys.* (N.Y.) **22** (1963), 1.

[5] M.-A. Tonnelat, *Ann. d. Phys.* **17** (1942), 159.

[6] M.-A. Tonnelat, in *Proceedings on Theory of Gravitation: Conference in Warszawa and Jabłonna*, 25-31 July, 1962, *ed.* L. Infeld. Gauthier-Villars, Paris (1964). This work refers to more of her papers.

[7] H. Weyl, *Am. J. Math.* **66** (1944), 591.

[8] A. Papapetrou, *Proc. Roy. Irish Acad. A* **52** (1948), 11.

[9] A. Papapetrou, W. Urich, *Annalen der Physik* **6** (1954), 220.

[10] M. Moshinsky, *Phys. Rev.* **80** (1950), 514.

[11] M. Kohler, *Zeit. Phys.* **131** (1952), 571.

[12] M. Kohler, *Zeit. Phys.* **134** (1953), 286, 306.

[13] S. N. Gupta, *Proc. Phys. Soc. (Lon.) A* **65** (1952), 161 and 608.

[14] S. N. Gupta, *Phys. Rev.* **96** (1954), 1683.

[15] S. N. Gupta, *Rev. Mod. Phys.* **29** (1957), 334.

[16] S. N. Gupta, *Phys. Rev.* **172** (1968), 1303.

[17] R. Kraichnan, *Phys. Rev.* **98** (1955), 1118.

[18] R. Kraichnan, *Phys. Rev.* **101** (1956), 482.

[19] F. J. Belinfante, *Phys. Rev.* **98** (1955), 793.

[20] F. J. Belinfante, J. C. Swihart, *Ann. Phys.* (NY) **1** (1957), 168, 196; **2** (1957) 81.

[21] F. J. Belinfante, J. C. Swihart, *Ann. Phys.* (NY) **2** (1957) 81.

[22] F. J. Belinfante, *Rev. Mex. Fis.* 4 (1955), 192.

[23] F. J. Belinfante, D. I. Caplan, W. L. Kennedy, *Rev. Mod. Phys.* **29** (1957), 518.

[24] F. J. Belinfante, J. C. Garrison, *Phys. Rev.* **125** (1962), 1124.

[25] W. Thirring, *Fortsch. Phys.* **7** (1959), 79.

[26] W. Thirring, *Ann. Phys. (N.Y.)* **16** (1961), 96.

[27] W. Thirring, *Acta Phys. Aust.* **suppl. 5** (1968), 391.

[28] W. Thirring, in *Contemporary Physics: Trieste Symposium 1968*, International Atomic Energy Agency, Vienna (1968).

[29] I. I. Gutman, *Sov. Phys. J. E. T. P.* **37** (1960), 1162.

[30] R. Sexl, in *Evidence for Gravitational Theories: Proceedings of the International School of Physics "Enrico Fermi"*, June 19-July 1, 1961, Varenna on Lake Como, Villa Monastero, *ed.* C. Møller. Academic, New York (1962).

[31] R. U. Sexl, *Fortschr. Phys.* **15** (1967), 269.

[32] O. Nachtmann, H. Schmidle, R. U. Sexl, *Acta Phys. Austr.* **29** (1968), 289.

[33] O. Nachtmann, H. Schmidle, R. U. Sexl, *Commun. Math. Phys.* **13** (1969), 254.

[34] S. Mandelstam, *Ann. Phys. (N.Y.)* **19** (1962), 25.

[35] D. E. Burlankov, *Sov. Phys. J. E. T. P.* **17** (1963), 1306.

[36] L. Halpern, *Bull. Cl. Sci. Acad. R. Belg., 5e serie,* **49** (1963), 226.

[37] L. Halpern, *Ann. Phys. (N.Y.)* **25** (1963), 387.

[38] J. R. Klauder, *Nuov. Cim.* **25** (1962), 542.

[39] E. R. Huggins, *Quantum Mechanics of the Interaction of Gravity with Electrons: Theory of a Spin-Two Field Coupled to Energy.* Dissertation, California Institute of Technology (1962).

[40] R. Feynman, F. Morinigo, W. Wagner, B. Hatfield, *Feynman Lectures on Gravitation*, Addison-Wesley, Reading, Mass. (1995).

[41] R. Feynman, *Acta Phys. Polon.* **24** (1963), 697.

[42] R. P. Feynman, in *Magic Without Magic: John Archibald Wheeler, A Collection of Essays in Honor of His Sixtieth Birthday*, *ed.* J. R. Klauder, Freeman, San Francisco (1972).

[43] R. Arnowitt, in *Proceedings of the Eastern Theoretical Physics Conference,* October 26-7, 1962, *ed.* M. E. Rose. Gordon and Breach, New York (1963).

[44] F. H. J. Cornish, *Proc. Roy. Soc.* **A282** (1964) 358, 372.

[45] S. Weinberg, *Phys. Rev.* **133** (1964), B1318.

[46] S. Weinberg, *Phys. Rev.* **134** (1964), B882.

[47] S. Weinberg, *Phys. Lett.* **9** (1964), 357.

[48] S. Weinberg, *Phys. Rev.* **135** (1964), B1049.

[49] S. Weinberg, "The Quantum Theory of Massless Particles," in *Lectures on Particles and Field Theory, 1964, volume 2, ed.* S. Deser and K. W. Ford, Prentice Hall, Englewood Cliffs, N. J. (1965).

[50] S. Weinberg, *Phys. Rev.* **138** (1965), B988.

[51] S. Weinberg, in *Contemporary Physics: Trieste Symposium 1968*, International Atomic Energy Agency, Vienna (1968).

[52] S. Weinberg, *Gravitation and Cosmology.* Wiley, New York (1972).

[53] S. Weinberg, E. Witten, *Phys. Lett. B* **96** (1980), 59.

[54] R. Penrose, in *Quantum Structure of Space and Time: Proceedings of the Nuffield Workshop, Imperial College London*, 3-21 August 1981, *ed.* M. J. Duff, C. J. Isham, Cambridge Univ., Cambridge (1981).

[55] W. Wyss, *Helv. Phys. Acta* **38** (1965), 469.

[56] W. Wyss, *Helv. Phys. Acta* **70** (1997), 336.

[57] V. I. Ogievetsky, I. V. Polubarinov, *Ann. Phys. (NY)* **35** (1965), 167.

[58] V. I. Ogievetskii, I. V. Polubarinov, *Sov. Phys. Dokl.* **11** (1966), 71.

[59] V. I. Ogievetskii, I. V. Polubarinov, *Sov. Phys. Dokl.* **11** (1966), 135.

[60] J. L. Anderson, *Principles of Relativity Physics.* Academic, New York (1967).

[61] P. Mittelstaedt, J. B. Barbour, *Z. Phys.* (1967), 82.

[62] P. Mittelstaedt, *Z. Phys.* **211** (1968), 271.

[63] P. Mittelstaedt, *Philosophical Problems of Modern Physics.* Reidel, Dordrecht (1976).

[64] P. Mittelstaedt, in *Einstein Symposium Berlin aus Anlass der 100. Wiederkehr seines Geburtstages, 25. bis 30. März 1979*, *ed.* H. Nelkowski, A. Hermann, H. Poser, R. Schrader, R. Seiler. Springer, Berlin (1979).

[65] K. Westpfahl, *Fortschr. Phys.* **15** (1967), 309.

[66] S. Deser, *Gen. Rel. Grav.* **1** (1970), 9.

[67] D. G. Boulware, S. Deser, *Ann. Phys. (N.Y.)* **89** (1975), 193.

[68] S. Deser, in *Quantum Gravity: An Oxford Symposium*, *ed.* C. J. Isham, R. Penrose, and D. W. Sciama, Clarendon, Oxford (1975).

[69] D. G. Boulware, S. Deser, J. H. Kay, *Physica A* **96** (1979), 141.

[70] S. Deser, *Class. Quant. Grav.* **4** (1987), L99.

[71] S. Deser, R. I. Nepomechie, *Phys. Let. B* **132** (1983), 321.

[72] S. Deser, R. I. Nepomechie, *Ann. Phys. (NY)* **154** (1984), 396.

[73] S. Deser, in *Developments in General Relativity, Astrophysics, and Quantum Theory: A Jubilee Volume in Honour of Nathan Rosen*, *ed.* F. I. Cooperstock, L. P. Horwitz, J. Rosen. I. O. P., Bristol, and the Israel Physical Society, Jerusalem (1990).

[74] S. Deser, A. Waldron, V. Pascalutsa, *Phys. Rev. D* **62** (2000), 105031.

[75] P. van Nieuwenhuizen, *Nuc. Phys. B* **60** (1973), 478.

[76] P. van Nieuwenhuizen, in *Proceedings of the First Marcel Grossmann Meeting on General Relativity*, 1975, Trieste, *ed.* R. Ruffini. North Holland, Amsterdam (1977).

[77] C. Rovelli, *Los Alamos Preprints*, www.arxiv.org, gr-qc/0006061; presented at the 9th Marcel Grossmann Meeting in Rome, July 2000.

[78] H. J. Groenewold, *Int. J. Theor. Phys.* **9** (1974), 19.

[79] J. Fang, C. Fronsdal, *J. Math. Phys.* **20** (1979), 2264.

[80] G. Cavalleri, G. Spinelli, *Rivista del Nuovo Cimento* **3** (1980), 1.

[81] R. Penrose, in *Essays in General Relativity—A Festschrift for Abraham Taub*, *ed.* F. J. Tipler. New York, Academic (1980).

[82] A. Mészáros, *Act. Phys. Hung.* **59** (1986), 379.

[83] L. P. Grishcuhk, A. N. Petrov, A. D. Popova, *Commun. Math. Phys.* **94** (1984), 379.

[84] Ya. B. Zel'dovich, L. P. Grishchuk, *Sov. Phys. Usp.* **29** (1986), 780.

[85] Ya. B. Zel'dovich, L. P. Grishchuk, *Sov. Phys. Usp.* **31** (1988), 666.

[86] L. P. Grishchuk, *Sov. Phys. Usp.* **33** (1990), 669.

[87] D. E. Burlankov, *Sov. J. Nucl. Phys.* **50** (1989), 174.

[88] R. M. Avakian, L. Sh. Grigorian, *Astro. Space Sci.* **146** (1988), 183.

[89] N. Straumann, *Los Alamos Preprints*, www.arxiv.org, astro-ph/0006423.

[90] A. A. Logunov, M. A. Mestvirishvili, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **86** (1991), 3.

[91] A. A. Logunov, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **92** (1992), 191.

[92] A. A. Logunov, *The Relativistic Theory of Gravitation.* Nova Science, Commack, New York (1998).

[93] I. Gottlieb, N. Ionescu-Pallas, *Rev. Roum. Phys.* **35** (1990), 25.

[94] S. Babak, L. Grishchuk, *Phys. Rev. D* **61** (2000), 024038.

[95] A. N. Petrov, J. Katz, *Los Alamos Preprints*, www.arxiv.org, gr-qc/9905088; to appear in the Proceedings of the Conference on "Fundamental Interactions: From Symmetries to Black Holes" held in Brussels March 25 - 27, 1999.

[96] A. N. Petrov, J. Katz, *Los Alamos Preprints*, www.arxiv.org, gr-qc/9911025.

[97] A. A. Vlasov, *Sov. J. Nucl. Phys.* **51** (1990), 1148; trans. *Yad. Fiz.*

[98] A. N. Petrov, *Astron. Astrophys. Trans.* **1** (1992), 195.

[99] Yu. M. Loskutov, *Sov. J. Nucl. Phys.* **54** (1991), 903.

[100] A. D. Popova, A. N. Petrov, *Int. J. Mod. Phys. A* **11** (1988), 2651,

[101] L. P. Grishchuk, A. N. Petrov, *Sov. Phys. J. E.T. P.* **65** (1987), 5; trans. *Zh. Eksp. Teor. Fiz.* **92** (1987), 9.

[102] Ya. I. Pugachev, V. D. Gun'ko, *Sov. Phys. Journ.*, trans. *Izvestiya Vysshikh Uchebnykh Zavednii, Fizika* **10** (1974), 46.

[103] A. N. Petrov, J. V. Narlikar, *Found. Phys.* **26**, 1201, 1996.

[104] A. N. Petrov, *Int. J. Mod. Phys. D* **4** (1995), 451.

[105] A. N. Petrov, *Int. J. Mod. Phys. D* **6** (1997), 239.

[106] A. A. Vlasov, *Moscow Unversity Physics Bulletin*, trans. *Vestnik Moskovskogo Universiteta. Fizika* **43** (1988), 87.

[107] E. Alvarez, *Rev. Mod. Phys.* **61** (1989), 561.

[108] Yu. V. Baryshev, *Gravitation* **2** (1996), 69; *Los Alamos Preprints*, www.arxiv.org, gr-qc/9912003.

[109] H. Nikolić, *Gen. Rel. Grav.* **31** (1999), 1211.

[110] H. Nikolić, *Los Alamos Preprints*, www.arxiv.org, gr-qc/9901045.

[111] A. I. Nikishov, *Los Alamos Preprrints*, www.arxiv.org, gr-qc/9912034.

[112] J. B. Pitts and W. C. Schieve, *Gen. Rel. Grav.* **33** (2001), forthcoming; *Los Alamos Preprints*, www.arxiv.org, gr-qc/0101058.

[113] G. Magnano, L. M. Sokołowski, *Los Alamos Preprints*, www.arxiv.org, gr-qc/0103072.

[114] G. Scharf, *Quantum Gauge Theories: A True Ghost Story.* Wiley, New York (2001).

[115] T. Borne, G. Lochak, H. Stumpf *Nonperturbative Quantum Field Theory and the Structure of Matter.* Kluwer, Dordrecht (2001).

[116] H. Fuchs, U. Kasper, D.-E. Liebscher, V. Müller, H.-J. Schmidt, *Fortschr. Phys.* **36** (1988), 427.

[117] G. Scharf, M. Wellmann, *Gen. Rel. Grav.* **33** (2001), 553.

[118] R. M. Avakian, L. Sh. Grigorian, *Astrophys. Space Sci.* **146** (1988), 183.

[119] J. B. Pitts, W. C. Schieve, in *Proceedings of the 20th Texas Symposium on Relativistic Astrophysics*, *ed.* H. Martel, J. C. Wheeler, forthcoming; gr-qc/0101097.

[120] J. Katz, A. Ori, *Class. Quant. Grav.* **7** (1990), 787.

[121] B. S. DeWitt, in *Gravitation: An Introduction to Current Research*, *ed.* L. Witten. Wiley, New York (1962).

[122] B. S. DeWitt, *Phys. Rev.* **162** (1967), 1195.

[123] B. S. DeWitt, in *Magic Without Magic: John Archibald Wheeler, A Collection of Essays in Honor of His Sixtieth Birthday, ed.* J. R. Klauder. Freeman, San Franciso (1972).

[124] T. A. Barnabey, *Phys. Rev. D* **10** (1974), 1741.

[125] L. Fatibene, M. Ferraris, M. Francaviglia, M. Raiteri, J. Math. Phys. **42** (2001), 1173; *Los Alamos Preprints*, www.arxiv.org, gr-qc/0003019.

[126] N. Straumann, *General Relativity and Relativistic Astrophysics.* Springer, Berlin (1984).

[127] A. Ashtekar, R. Geroch, *Rep. Prog. Phys.* **37** (1974), 1211.

[128] R. Geroch, *J. Math. Phys.* **9** (1968), 1739.

[129] S. W. Hawking, G. F. R. Ellis, *The Large Scale Structure of Space-time.* Cambridge Univ., Cambridge (1973).

[130] R. Wald, *Phys. Rev. D* **33** (1986), 3613.

[131] K. R. Heiderich, W. G. Unruh, *Phys. Rev. D* **38** (1988), 490; also *Phys. Rev. D* **42** (1990), 2057.

[132] L. P. Grishchuk, A. D. Popova, *Sov. Phys. J. E. T. P.* **53** (1981), 1.

[133] L. P. Grishchuk, A. D. Popova, *J. Phys. A.: Math. Gen.* **15** (1982), 3525.

[134] C. Misner, K. Thorne, and J. Wheeler, *Gravitation.*. Freeman, New York (1973).

[135] R. Wald, *General Relativity.* Univ. Chicago, Chicago (1984).

[136] C. J. Isham, Lectures presented at the NATO Advanced Study Institute "Recent Problems in Mathematical Physics," Salamanca, June 15-27, 1992. *Los Alamos Preprints*, www.arxiv.org, gr-qc/9210011.

[137] W. L. Craig, *Time and the Metaphysics of Relativity*, Kluwer Academic, Dordrecht (2001).

[138] E. Prugovečki, *Quantum Geometry: A Framework for Quantum General Relativity.* Kluwer Academic, Dordrecht (1992).

[139] J. Butterfield, C. J. Isham, in *Physics Meets Philosophy at the Planck Scale, ed.* C. Callender, N. Huggett. Cambridge Univ., Cambridge (2001); *Los Alamos Preprints*, www.arxiv.org, gr-qc/9903072.

[140] R. Adler, M. Bazin, M. Schiffer, *Introduction to General Relativity.* McGraw-Hill, New York (1965).

[141] H. Stephani, *General Relativity*, 2nd ed. Cambridge Univ., Cambridge (1990).

[142] W. Pauli, *Theory of Relativity, trans.* G. Field. Pergamon, New York (1958).

[143] C. Møller, *Mat. Fys. Skr. Dan. Vid. Selsk.* **1** (1961), no. 1.

[144] C. Møller, in *Evidence for Gravitational Theories: Proceedings of the International School of Physics "Enrico Fermi," course XX*, Varenna on Lake Como, Villa Monastero, June 19-July 1, 1961, *ed.* C. Møller. Academic, New York (1963).

[145] C. Møller, in *Proceedings on Theory of Gravitation: Conference in Warszawa and Jabłonna*, 25-31 July, 1962, *ed.* L. Infeld. Gauthier-Villars, Paris (1964).

[146] C. Møller, *Mat. Fys. Medd. Dan. Vid. Selsk.* **34** (1964), no. 3.

[147] C. Møller, *Mat. Fys. Medd. Dan. Vid. Selsk.* **35** (1966), no. 3.

[148] G. Y. Chee, Y. Zhang, Y. Guo, *Los Alamos Preprints*, www.arxiv.org, gr-qc/0106053.

[149] N. Rosen, *Found. Phys.* **15** (1985), 997.

[150] N. Rosen, in *Proceedings on Theory of Gravitation: Conference in Warszawa and Jabłonna*, 25-31 July, 1962, *ed.* L. Infeld. Gauthier-Villars, Paris (1964).

[151] B. A. Bassett, S. Liberati, C. Molina-Paris, M. Visser, *Los Alamos Preprints*, www.arxiv.org, astro-ph/0001441.

[152] G. Velo, Z. Zwanziger, *Phys. Rev.* **186** (1969), 1337.

[153] D. Boulware, S. Deser, *Phys. Rev. D* **6** (1972), 3368.

[154] M. A. Clayton, *Class. Quantum Grav.* **13** (1996), 2851.

[155] S. Deser, *Rev. Mod. Phys.* **29** (1957), 417.

[156] L. P. Grishchuk, A. N. Petrov, *Sov. Astron. Lett.* **12** (1986), 179, trans. *Pis'ma Astron. Zh.* **12** (1986), 439.

[157] A. A. Vlasov, *Moscow Uni. Phys. Bull.* **43** (1988), 72, trans. *Vestnik Moskovskogo Universiteta. Fizika.*

[158] Yu. P. Viblyi, *Theoretical and Mathematical Physics*, trans. *Teoeticheskaya i Matematicheskaya Fizika* **88** (1991), 135.

[159] J. L. Anderson, *Los Alamos Preprints*, www.arxiv.org, gr-qc/9912051.

[160] S. Deser, C. J. Isham, *Phys. Rev. D* **14** (1976), 2505.

[161] J. M. Charap, *Class. Quant. Grav.* **4** (1987), 247.

[162] J. Bičák, in *Einstein Field Equations and Their Physical Implications–Selected essays in honour of Juergen Ehlers, ed.* B. G. Schmidt, Springer, Berlin (2000); *Los Alamos Preprints*, www.arxiv.org, gr-qc/0004016

[163] Yu. M. Loskutov, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **94** (1993), 515.

[164] Yu. V. Chugreev, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **88** (1991), 459.

[165] K. H. E. Pinson, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **89** (1991), 317.

[166] H. Goldstein, *Classical Mechanics, 2nd ed.* Addison-Wesley, Reading, Mass. (1980).

[167] C. J. Isham, in *Quantum Gravity: An Oxford Symposium*, *ed.* C. J. Isham, R. Penrose, and D. W. Sciama, Clarendon, Oxford (1975).

[168] C. J. Isham, in *Quantum Gravity 2: A Second Oxford Symposium*, *ed.* C. J. Isham, R. Penrose, and D. W. Sciama, Clarendon, Oxford (1981).

[169] S. Weinstein, in *Physics Meets Philosophy at the Planck Scale*, *ed.* C. Callender, N. Huggett. Cambridge Univ., Cambridge (2001).

[170] J. R. Weeks, *Class. Quant. Grav.* **15** (1998), 2599; *Los Alamos Preprints*, www.arxiv.org, astro-ph/9802012.

[171] E. Gausmann, R. Lehoucq, J.-P. Luminet, J.-P. Uzan, J. Weeks, *Los Alamos Preprints*, www.arxiv.org, gr-qc/0106033.

[172] V. Blanloeil, B. F. Roukema, *Los Alamos Preprints*, www.arxiv.org, astro-ph/0010170.

[173] G.I. Gomero, M.J. Reboucas, R. Tavakol, *Class. Quant. Grav.* **18** (2001), forthcoming; *Los Alamos Preprints*, www.arxiv.org, gr-qc/0105002.

[174] B. F. Roukema, *Bull. Astron. Soc. India* **28** (2000), 483; *Los Alamos Preprints*, www.arxiv.org, astro-ph/0010185.

[175] J. Levin, *Los Alamos Preprints*, www.arxiv.org, gr-qc/0108043.

[176] J. A. Schouten, *Ricci-Calculus: An Introduction to Tensor Analysis and Its Geometrical Applications, 2nd ed.* Springer, Berlin (1954).

[177] Y. Choquet-Bruhat, S. Deser, *Ann. Phys. (NY)* **81** (1973), 165.

[178] D. R. Brill, S. Deser, *Comm. Math. Phys.* **32** (1973), 291.

[179] C. J. Isham, writeup of plenary lecture, GR14 conference, Florence, 1995; *Los Alamos Preprints*, www.arxiv.org, gr-qc/9510063.

[180] G. Belot, J. Earman, L. Ruetsche, *Brit. J. Phil. Sci.* **50** (1999), 189.

[181] J. Oppenheim, *Los Alamos Preprints*, www.arxiv.org, gr-qc/0105101.

[182] H. Ohanian, R. Ruffini, *Gravitation and Spacetime*, 2nd ed. Norton, New York (1994).

[183] J. T. Oden, J. N. Reddy, *Variational Methods in Theoretical Mechanics.* Springer, Berlin (1976).

[184] G. Del Piero, F. Maceri, *ed.*, *Unilateral Problems in Structural Analysis: Proceedings of the Second [*sic*] Meeting on Unilateral Problems in Structural Analysis*, 1983, Ravello, *ed.* International Centre for Mechanical Science # 288, Springer, Vienna (1985).

[185] G. Del Piero, F. Maceri, *ed.*, *Unilateral Problems in Structural Analysis - 2: Proceedings of the Second [*sic*] Meeting on Unilateral Problems in Structural Analysis*, 1985, Prescudin, International Centre for Mechanical Science # 304, Springer, Vienna (1987).

[186] J. J. Moreau, P. D. Panagiotopoulos, G. Strang, *ed.*, *Topics in Nonsmooth Mechanics.* Birkhäuser, Basel (1988).

[187] M. D. P. Monteiro Marques, *Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction*, Birkhäuser, Basel (1993).

[188] E. A. Lacomba, W. M. Tulczyjew, *J. Math. Phys.* **23** (1990), 2801.

[189] A. Ibort, M. de León, E. A. Lacomba, D. M. de Diego, P. Pitanga, *J. Phys. A.* **30** (1997), 5835.

[190] F. Cantrijn, M. de León, J. C. Marrero, D. M. de Diego, *J. Math. Phys.* **40** (1999), 795.

[191] F. S. Klotz, *A Study of General Relativity in Canonical Form.* Dissertation, Syracuse University (1970).

[192] J. N. Goldberg, F. S. Klotz, *Int. J. Theor. Phys.* **7** (1973), 31.

[193] K. Sundermeyer, *Constrained dynamics.* Springer, Berlin (1982).

[194] P. A. M. Dirac, *Proc. Roy. Soc. (London) A* **246** (1958), 333.

[195] J. L. Anderson, *Phys. Rev.* **111** (1958), 965.

[196] R. T. Jantzen, in *Proc. Int. Sch. Phys. "E. Fermi" Course LXXXVI (1982) on "Gamov Cosmology," ed.* R. Ruffini, F. Melchiorri. North Holland, Amsterdam (1987); *Los Alamos Preprints*, www.arxiv.org, gr-qc/0001069.

[197] J. M. Pons, D. C. Salisbury, L. C. Shepley, *J. Phys. A: Math. Gen.* **32**, 419 (1999).

[198] G. S. Hall, *Arab. J. Sci. Eng.* **9** (1984), 87.

[199] G. S. Hall, in *Differential Geometry*, Banach Center, PWN–Polish Scientific Publishers, Warsaw (1984).

[200] G. S. Hall, D. A. Negm, *Int. J. Theor. Phys.* **25** (1986), 405.

[201] G. S. Hall, M. J. Rebouças, J. Santos, A. F. F. Teixeira, *Gen. Rel. Grav.* **28** (1996), 1107.

[202] C. Bona, B. Coll, J. A. Morales, *J. Math. Phys.* **33** (1992), 670.

[203] I. Goldman, *Gen. Rel. Grav.* **9** (1978), 575.

[204] K. P. Tod, *Class. Quantum. Grav.* **9** (1992), 1693.

[205] R. V. Churchill, *Trans. Amer. Math. Soc.* **34** (1932), 784.

[206] M. Kaku, *Quantum Field Theory: A Modern Introduction.* Oxford, New York (1993).

[207] C. J. Isham, A. C. Kakas, *Class. Quantum Grav.* **1** (1984), 621, 633.

[208] K. V. Kuchař, in *General Relativity and Gravitation 1992: Proceedings of the Thirteenth International Conference on General Relativity and Gravitation held at Cordoba, Argentina, 28 June - 4 July 1992*, *ed.* R. J. Gleiser, C. N. Kozameh, O. M. Moreschi. Institute of Physics, Bristol (1993).

[209] J. R. Klauder, *Los Alamos Preprints*, www.arxiv.org, gr-qc/0102041.

[210] B. Coll, F. San José Martínez, *Los Alamos Preprints*, www.arxiv.org, gr-qc/01006052.

[211] O. Lauscher, M. Reuter, *Los Alamos Preprints*, www.arxiv.org, hep-th/0108040.

[212] R. V. Churchill, J. W. Brown, *Complex Variables and Applicaitons*, (5th ed.). McGraw-Hill, New York (1990).

[213] Y. Choquet-Bruhat, C. DeWitt-Morette, M. Dillard-Bleick, *Analysis, Manifolds, and Physics. Part I: Basics*, rev. ed. North-Holland, Amsterdam (1996).

[214] J. D. Jackson, *Classical Electrodynamics*, 2nd. ed. Wiley, New York (1975).

[215] A. Anderson, Y. Choquet-Bruhat, J. W. York Jr., to be published in the Proceedings of the 2nd Samos Meeting; *Los Alamos Preprints*, www.arxiv.org, gr-qc/9907099.

[216] C. J. Isham, in *Canonical Gravity: From Classical to Quantum. Proceedings, of the 117th WE Heraeus Seminar Held at Bad Honnef, Germany, 13-17 September 1993, ed.* J. Ehlers, H. Friedrich. Springer, Berlin (1994); *Los Alamos Preprints*, www.arxiv.org, gr-qc/9310031.

[217] A. A. Logunov, M. A. Mestvirishvili, *Los Alamos Preprints*, www.arxiv.org, gr-qc/0106055.